



TUGAS AKHIR - SS141501

**Deteksi Dini Osteoporosis pada Wanita *Postmenopausal*
Melalui *Dental Panoramic Radiograph* Menggunakan
Metode Regresi Logistik Biner, *Classification Tree* dan
*Support Vector Machine***

**M. RIZKY FAUZY
NRP 1312 100 053**

**Dosen Pembimbing
Dr. Kartika Fithriasari, M.Si
Dr. Agus Zainal Arifin, S.Kom, M.Kom**

**Program Studi S1
Jurusan Statistika
Fakultas Matematika Dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2016**



FINAL PROJECT - SS141501

Early Detection of Osteoporosis in Postmenopausal Women by Dental Panoramic Radiograph Using Binary Logistic Regression, Classification Tree and Support Vector Machine

M. RIZKY FAUZY
NRP. 1312 100 053

Supervisor
Dr. Kartika Fithriasari, M.Si
Dr. Agus Zainal Arifin, S.Kom, M.Kom

Undergraduate Programme
Department of Statistics
Faculty of Mathematics and Natural Sciences
Institut Teknologi Sepuluh Nopember
Surabaya 2016

LEMBAR PENGESAHAN

**DETEKSI DINI OSTEOPOROSIS PADA WANITA
POSTMENOPAUSAL MELALUI DENTAL PANORAMIC
RADIOGRAPH MENGGUNAKAN METODE REGRESI
LOGISTIK BINER, CLASSIFICATION TREE
DAN SUPPORT VECTOR MACHINE**

TUGAS AKHIR

Diajukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi S-1 Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :

**M. RIZKY FAUZY
NRP. 1312 100 053**

Disetujui oleh Pembimbing Tugas Akhir

Dr. Kartika Fithriasari, M.Si

NIP. 19691212 199303 2 002

Dr. Agus Zainal Arifin, S.Kom, M.Kom

NIP. 19720809 199512 1 001

Mengetahui

Ketua Jurusan Statistika FMIPA-ITS

Dr. Suhartono

NIP. 19710929 199512 1 001

SURABAYA, JULI 2016



**DETEKSI DINI OSTEOPOROSIS PADA WANITA
POSTMENOPAUSAL MELALUI DENTAL PANORAMIC
RADIOGRAPH MENGGUNAKAN METODE REGRESI
LOGISTIK BINER, CLASSIFICATION TREE
DAN SUPPORT VECTOR MACHINE**

Nama : M. Rizky Fauzy
NRP : 1312 100 053
Jurusan : Statistika FMIPA-ITS
Pembimbing : Dr. Kartika Fithriasari, M.Si
Dr. Agus Zainal Arifin, S.Kom, M.Kom

Abstrak

Osteoporosis adalah berkurangnya massa tulang dan perubahan arsitekturnya sampai batas patah tanpa keluhan. Risiko wanita terkena penyakit ini lebih besar daripada pria. Deteksi dini penyakit ini perlu dilakukan agar berbagai dampak buruk yang mungkin terjadi dapat diminimalisir. Namun harga alat deteksi osteoporosis, DXA, sangat mahal dan ketersediaannya masih sangat terbatas, bahkan di negara maju sekalipun. Telah banyak penelitian yang menunjukkan bahwa terdapat beberapa informasi yang bisa diambil dari Dental Panoramic Radiograph (DPR) sebagai upaya deteksi osteoporosis, seperti lebar cortical bone dan kekuatan garis pada trabecular bone. Namun deteksi yang dilakukan hanya menggunakan salah satu informasi saja, dan hanya menggunakan suatu nilai threshold. Oleh karena itu, pada tugas akhir ini akan digunakan kedua informasi tersebut sebagai upaya deteksi dini osteoporosis. Metode klasifikasi yang digunakan adalah metode Regresi Logistik Biner, Classification Tree (CTree) dan Support Vector Machine (SVM). Citra yang digunakan merupakan DPR dari 100 wanita postmenopausal yang berusia lebih dari 50 tahun. Hasil penelitian menunjukkan bahwa metode SVM memiliki performa yang lebih baik daripada metode lainnya, dimana accuracy, sensitivity, dan specificity yang didapat untuk data testing adalah 77,00%, 70,00%, dan 85,87%.

Kata Kunci : *Classification Tree, Dental Panoramic Radiograph, Osteoporosis, Regresi Logistik Biner, Support Vector Machine, Wanita Postmenopausal*

EARLY DETECTION OF OSTEOPOROSIS IN POSTMENOPAUSAL WOMEN BY DENTAL PANORAMIC RADIOGRAPH USING BINARY LOGISTIC REGRESSION, CLASSIFICATION TREE AND SUPPORT VECTOR MACHINE

Name : M. Rizky Fauzy
NRP : 1312 100 053
Department : Statistics FMIPA-ITS
Supervisor : Dr. Kartika Fithriasari, M.Si
Dr. Agus Zainal Arifin, S.Kom, M.Kom

Abstract

Osteoporosis is a condition where bone mass is decreasing and change of its architecture until fracture without any indication. This disease often known as silent disease. Women more prone to osteoporosis than men. Early detection of osteoporosis is needed to prevent this bone fracture happen. But, the common tools for osteoporosis detection, DXA, is very expensive and not every hospitals has DXA, even in developed countries. On the other hand, there are many research that use information that can be retrieve from Dental Panoramic Radiograph (DPR) for osteoporosis detection, such as cortical bone width and trabecular bone line strength. But the detection only consider one information only, and the classification process is done using a threshold. So that this final project propose an early detection of osteoporosis using both information at the same time. Furthermore the classification process is done using classification methods such as Binary Logistic Regression, Classification Tree (CTree) and Support Vector Machine (SVM). The images that used in this research are DPR from 100 postmenopausal womens aged 50 years or more. This research show that SVM method is better than other method, where the overall accuracy, sensitivity, and specificity using SVM method with K-folds cross validation is 77,00%, 70,00%, and 85,87% respectively.

Keywords : Binary Logistic Regression, Classification Tree, Dental Panoramic Radiograph, Osteoporosis, Postmenopausal Women, Support Vector Machine

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
TITLE PAGE	iii
LEMBAR PENGESAHAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	6
1.4 Manfaat	6
1.5 Batasan Masalah	6
BAB II TINJAUAN PUSTAKA	
2.1 Osteoporosis	9
2.2 <i>Dental Panoramic Radiograph</i>	9
2.3 Struktur Tulang pada Rahang	9
2.4 <i>Image Processing</i>	11
2.5 <i>Line Strength Detection</i>	12
2.6 <i>Cortical Bone Width Measurement</i>	13
2.7 <i>K-folds Cross Validation (KCV)</i>	15
2.8 Regresi Logistik Biner	17
2.8.1 Pengecekan Asumsi Multikolinieritas	18
2.8.2 Estimasi Parameter	18
2.8.3 Uji Signifikansi Parameter	22

2.9 <i>Classification Tree</i> (CTree)	23
2.10 <i>Support Vector Machine</i> (SVM)	25
2.10.1 Kasus Data yang Bisa Dipisah Secara Linier	26
2.10.2 Kasus Data yang Tidak Bisa Dipisah Secara Linier	29
2.10.3 Fungsi Kernel pada SVM untuk Kasus Data Non- Linier	30
2.11 Penelitian Sebelumnya	32
BAB III METODOLOGI PENELITIAN	
3.1 Sumber Data	33
3.2 Variabel Penelitian	33
3.3 Langkah Analisis	34
3.4 Diagram Alir	36
BAB IV ANALISIS DAN PEMBAHASAN	
4.1 Deskripsi Data Hasil <i>Preprocessing</i> Citra Rontgen Gigi	37
4.2 Klasifikasi BMD pada Wanita <i>Postmenopausal</i>	40
4.2.1 Klasifikasi dengan Metode Regresi Logistik Biner	40
4.2.2 Klasifikasi dengan Metode <i>Classification Tree</i>	43
4.2.3 Klasifikasi dengan Metode <i>Support Vector</i> <i>Machine</i>	48
4.2.4 Perbandingan Performa Antar Metode	50
4.3 <i>Graphic User Interface</i> (GUI) Deteksi Dini Osteoporosis dari Citra Rontgen Gigi	51
BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan	59
5.2 Saran	59
DAFTAR PUSTAKA	61
LAMPIRAN	65

DAFTAR TABEL

	Halaman
Tabel 2.1 Beberapa Fungsi Kernel yang Umum Digunakan Pada SVM	32
Tabel 3.1 Variabel yang Digunakan dalam Penelitian	33
Tabel 4.1 Nilai VIF Setiap Variabel Prediktor	40
Tabel 4.2 Performa Metode Regresi Logistik Biner Menggunakan Semua Variabel	41
Tabel 4.3 Nilai Koefisien, statistik hitung, dan <i>pvalue</i> dari Model pada <i>Fold</i> ke-7 untuk Semua Variabel Masuk dalam Model	41
Tabel 4.4 Nilai Koefisien, statistik hitung, dan <i>pvalue</i> Hasil <i>Backward Elimination</i>	42
Tabel 4.5 Performa Metode Regresi Logistik Biner Menggunakan Variabel RCB dan LTB	42
Tabel 4.6 Nilai Koefisien, statistik hitung, dan <i>pvalue</i> dari Model Pada <i>Fold</i> ke-7 dengan Variabel Prediktor RCB dan LTB	42
Tabel 4.7 Performa Metode <i>Classification Tree</i> Menggunakan Semua Variabel	43
Tabel 4.8 Beberapa Kondisi Pengelompokkan Hasil Pohon Klasifikasi dengan Semua Variabel Masuk dalam Model	44
Tabel 4.9 Performa Metode <i>Classification Tree</i> Menggunakan Variabel RCB dan LTB	45
Tabel 4.10 Beberapa Kondisi Pengelompokkan Hasil Pohon Klasifikasi Menggunakan Variabel RCB dan LTB	47
Tabel 4.11 Rangkuman Hasil KCV dari Metode SVM untuk Setiap Fungsi Kernel	48
Tabel 4.12 Performa Metode <i>Support Vector Machine</i> Menggunakan Variabel RCB dan LTB	49

Tabel 4.13	Rangkuman Performa dari Metode Regresi Logistik Biner (RLB), <i>Classification Tree</i> (CTree), dan <i>Support Vector Machine</i> (SVM) ...	51
Tabel 4.14	Penjelasan Fungsi Tombol-tombol pada GUI	52

DAFTAR GAMBAR

	Halaman
Gambar 2.1 (a) Posisi Pengambilan <i>Panoramic Radiograph</i> , (b) Hasil <i>Panoramic Radiograph</i>	10
Gambar 2.2 Letak <i>Trabecular Bone</i> dan <i>Cortical Bone</i> pada Rahang Gigi	10
Gambar 2.3 Struktur <i>Trabecular Bone</i> dan <i>Cortical Bone</i> pada Tulang	10
Gambar 2.4 Ilustrasi Citra ROI untuk Rahang Sebelah (a) Kanan, (c) Kiri, dan Citra Hasil Pengolahan untuk Rahang Sebelah (b) Kanan, (d) Kiri	13
Gambar 2.5 Ilustrasi Proses Pengukuran Lebar <i>Cortical Bone</i> , Dimana (a), (b) Citra Hasil <i>Thresholding</i> , (c), (d) Citra Hasil <i>High-pass Filtering</i> , dan (e), (f) Citra Hasil Pengukuran Akhir pada Rahang Sebelah Kanan dan Rahang Sebelah Kiri	15
Gambar 2.6 Ilustrasi Pembagian Data Menggunakan KCV	16
Gambar 2.7 Contoh Pohon Klasifikasi pada Kasus Pembelian Komputer	23
Gambar 2.8 (a) Ilustrasi Beberapa <i>Hyperplane</i> yang Bisa Digunakan pada Data yang Bisa Dipisah Secara Linier, Dimana (b) <i>Hyperplane</i> yang Menghasilkan Margin yang Kecil, (c) <i>Hyperplane</i> yang Menghasilkan Margin yang Besar	26
Gambar 2.9 Ilustrasi Kasus Data yang Tidak Bisa Dipisah Secara Linier	29
Gambar 2.10 (a) Ilustrasi Kasus Data Non-Linier, (b) Hasil Transformasi Data Beserta <i>Hyperplane</i> yang Didapat	31

Gambar 3.1	Langkan-langkah yang Akan Dilakukan dalam Penelitian	36
Gambar 4.1	Ilustrasi ROI yang Dianalisis dalam <i>Preprocessing</i>	37
Gambar 4.2	Perbedaan Sebaran Data Antara Wanita dengan ‘BMD Normal’ dan Wanita dengan ‘BMD Rendah’ pada (a) <i>Cortical Bone</i> dan (b) <i>Trabecular Bone</i>	38
Gambar 4.3	Sebaran Data Antar Kategori ‘BMD Normal’ dan ‘BMD Rendah’ pada Setiap Variabel Prediktor	39
Gambar 4.4	Pohon Klasifikasi yang Didapat pada <i>Fold</i> ke-7 Menggunakan Semua Variabel Prediktor	44
Gambar 4.5	Pohon Klasifikasi yang Didapat pada <i>Fold</i> ke-2 Menggunakan Variabel RCB dan LTB	46
Gambar 4.6	Tampilan Menu Utama GUI pada Kondisi Awal	52
Gambar 4.7	Tampilan GUI Setelah Citra Rontgen Gigi Dipilih	53
Gambar 4.8	Tampilan GUI (a) Ketika Tombol ‘ <i>Trabecular Bone</i> ’ ditekan dan (b) Setelah Proses Perhitungan pada <i>Trabecular Bone</i> Selesai	54
Gambar 4.9	Tampilan yang Muncul Ketika Tombol ‘ <i>Cortical Bone</i> ’ Ditekan	55
Gambar 4.10	Tampilan GUI Ketika Proses pada Tombol ‘ <i>Cortical Bone</i> ’ Selesai	56
Gambar 4.11	Tampilan GUI Ketika Proses pada Tombol ‘ <i>Diagnosis</i> ’ Selesai	57

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut data yang dikeluarkan oleh *World Health Organization* (WHO), telah tercatat sebanyak 200 juta orang di dunia menderita osteoporosis. Di Indonesia, yang saat ini memiliki jumlah penduduk hampir 250 juta jiwa, diperkirakan 28,7% laki-laki dan 32,3% perempuan didiagnosis terkena osteoporosis (Tribun News, 2015). Jumlah penderita osteoporosis ini terus bertambah setiap tahunnya, dimana pada tahun 2012, tercatat 2 dari 5 orang penduduk Indonesia rentan terkena osteoporosis (Tribun News, 2012). Osteoporosis adalah berkurangnya massa tulang dan perubahan arsitekturnya sampai batas patah tanpa keluhan. Penyakit ini tidak menimbulkan keluhan apapun sampai tulang penderita patah atau retak (Pangkalan Ide, 2012). Hal ini tentu akan membatasi aktivitas dalam kehidupan sehari-hari, sehingga menyebabkan produktivitas menurun. Oleh karena itu, perlu adanya deteksi dini untuk mengetahui apakah seseorang menderita osteoporosis atau tidak, sehingga kejadian tulang retak atau bahkan tulang patah dapat dihindari.

Penderita osteoporosis kebanyakan terjadi pada orang dengan usia di atas 50 tahun, terutama berjenis kelamin wanita. Risiko osteoporosis terjadi mulai sekitar umur 40 tahun, dan sekitar umur 50 tahun risiko terkena osteoporosis semakin tinggi. Faktor risiko wanita terkena osteoporosis tiga kali lebih besar dibandingkan pria (WHO Scientific Group, 2003). Perkiraan global saat ini satu dari tiga wanita terkena osteoporosis, sedangkan untuk pria sekitar satu banding empat. Hal tersebut terjadi karena penurunan fungsi hormon estrogen pada wanita. Hormon ini mengendalikan seluruh sistem pada tubuh wanita. Penurunan fungsi estrogen juga dapat mengakibatkan menstruasi berhenti (*menopause*), rambut rontok, pelupa, cepat marah, dll (Pangkalan Ide, 2012). Oleh karena itu, perlu adanya perhatian khusus mengenai osteoporosis pada wanita *postmenopausal* agar berbagai dampak buruk yang mungkin terjadi dapat diminimalisir.

Pemeriksaan osteoporosis umumnya dilakukan dengan cara melakukan pengecekan kekuatan garis pada tulang atau *bone mineral density* (BMD). Telah terdapat beberapa alat yang dapat digunakan untuk mengukur BMD dan memprediksi risiko osteoporosis, salah satunya adalah menggunakan *Dual-energy X-ray Absorptiometry* (DXA). Namun harga DXA sangat mahal dan ketersediaan alat ini masih sangat terbatas, bahkan di negara maju sekalipun. Kemudian umumnya penderita yang terkena penyakit osteoporosis, khususnya wanita *postmenopausal*, tidak melakukan pemeriksaan BMD rutin ke rumah sakit atau laboratorium rumah sakit, karena penyakit ini tidak menimbulkan keluhan sama sekali. Mereka lebih sering melakukan pemeriksaan gigi atau pengobatan gigi. Oleh karena itu, banyak citra *Dental Panoramic Radiograph* (DPR) yang telah dihasilkan untuk pemeriksaan gigi (Arifin et al., 2010).

Telah banyak penelitian yang menunjukkan bahwa terdapat beberapa informasi yang dapat diambil dari rahang gigi dalam upaya deteksi osteoporosis. Salah satu informasi yang dapat digunakan adalah lebar *cortical bone*. Berdasarkan penelitian yang telah dilakukan oleh Taguchi et al. (2004) didapat kesimpulan bahwa deteksi dini osteoporosis yang dilakukan melalui *Dental Panoramic Radiograph* menghasilkan kesimpulan yang sama dengan deteksi osteoporosis yang dilakukan melalui *Osteoporosis Self-assessment Tools* (OST). Arifin et al. (2006) juga melakukan penelitian untuk mendeteksi osteoporosis pada wanita *postmenopausal* melalui lebar *cortical bone*. Penelitian ini menghasilkan kesimpulan bahwa terdapat korelasi antara lebar *cortical bone* dengan BMD, namun akurasi deteksi yang dihasilkan masih berada di bawah 70,5%. Wardhana (2009) juga melakukan penelitian untuk mengukur lebar *cortical bone* dengan menggunakan *Active Shape Models* (ASM), dimana hasil lebar *cortical bone* yang didapat menggunakan metode tersebut mendekati lebar *cortical bone* sebenarnya dengan akurasi lebih dari 90%, namun pada penelitian tersebut lebar *cortical bone* yang didapat tidak digunakan lebih lanjut untuk deteksi osteoporosis.

Selain lebar *cortical bone*, informasi lain yang dapat digunakan untuk deteksi osteoporosis adalah *trabecular bone*. Penelitian ini telah dilakukan oleh White dan Rudolph (1999) dan menghasilkan kesimpulan bahwa orang yang terkena osteoporosis memiliki struktur *trabecular bone* pada rahang yang berubah daripada orang normal. Perubahan tersebut dapat berupa pengurangan kekuatan, kepadatan, ataupun kompleksitas strukturnya. Penelitian tentang *trabecular bone* juga telah dilakukan oleh Arifin et al. (2010) yang melakukan deteksi osteoporosis dari kekuatan garis pada *trabecular bone*, dimana hasil akurasi yang didapat dalam mendeteksi pasien yang terkena osteoporosis adalah sebesar 88%. Wihandika (2013) juga melakukan penelitian mengenai *trabecular bone* dengan menghitung banyaknya percabangan yang terdapat pada *trabecular bone*. Akurasi yang didapat dalam mendeteksi banyak percabangan pada *trabecular bone* adalah 95,25%, namun ketepatan klasifikasi pasien yang menderita osteoporosis yang didapat hanya 55,10%.

Berbagai penelitian tersebut telah melakukan klasifikasi melalui *Dental Panoramic Radiograph* untuk mendeteksi apakah seorang pasien terkena osteoporosis atau tidak, namun klasifikasi yang dilakukan masih berdasarkan satu indikator atau variabel saja. Klasifikasi yang dilakukan juga masih berdasarkan suatu nilai *threshold* atau nilai batasan tertentu. Oleh karena itu, perlu adanya deteksi dini osteoporosis dengan menggunakan berbagai informasi yang bisa didapat dari DPR, serta menggunakan suatu metode klasifikasi, agar didapat akurasi yang lebih baik.

Pada penelitian ini akan dilakukan klasifikasi osteoporosis dengan menggabungkan berbagai indikator atau variabel yang pernah digunakan pada penelitian sebelumnya, yaitu lebar *cortical bone* dan kekuatan garis pada *trabecular bone*. Klasifikasi akan dilakukan dengan menggunakan metode *Support Vector Machine* (SVM). Metode ini digunakan karena mampu melakukan klasifikasi pada data yang bisa dipisahkan oleh suatu garis linier maupun tidak (data non-linier). Umumnya akurasi yang dihasilkan dari metode ini tinggi dan cenderung untuk tidak terjadi kasus *over-*

fitting. Menurut Auria dan Moro (2008) metode SVM memiliki beberapa kelebihan, beberapa diantaranya adalah SVM memiliki sifat global optima dan dapat digunakan untuk data non-linier, karena fungsi kernel pada SVM secara implisit memuat transformasi non-linier. Metode SVM pernah digunakan oleh Ogutu et al. (2011) untuk melakukan *Genome Selection* (GS) berdasarkan pembiakan menggunakan molekul yang ada di gen tertentu. Berdasarkan penelitian tersebut, diketahui bahwa metode SVM memiliki akurasi yang relatif lebih baik jika dibandingkan dengan metode *Random Forest*, selain itu SVM juga memiliki waktu komputasi yang lebih cepat dibandingkan dengan *Random Forest*. Penelitian lain tentang SVM dilakukan oleh Fachruddin (2015), yang melakukan perbandingan metode *Random Forest* dan SVM untuk deteksi epilepsi menggunakan data rekaman *electro-encephalograph*. Penelitian tersebut menghasilkan kesimpulan bahwa rata-rata akurasi data *testing* yang dihasilkan metode SVM lebih baik daripada metode *Random Forest*, namun untuk data *training*, rata-rata akurasi metode *Random Forest* lebih baik daripada metode SVM.

Selain melakukan klasifikasi, akan dibuat juga suatu model yang dapat menjelaskan hubungan antara variabel prediktor, yaitu lebar *cortical bone* dan kekuatan garis pada *trabecular bone*, dengan variabel respon, yaitu kategori apakah wanita tersebut memiliki BMD yang rendah atau BMD yang normal. Pembuatan model akan dilakukan dengan menggunakan metode Regresi Logistik, karena metode ini merupakan suatu metode klasifikasi parametrik yang rendah asumsi jika dibandingkan dengan metode parametrik lainnya. Asumsi yang perlu dipenuhi dalam metode ini adalah asumsi tidak terjadi multikolinieritas. Model yang didapat menggunakan metode ini juga dapat digunakan untuk melihat apakah variabel prediktor tersebut berpengaruh secara signifikan dalam klasifikasi pasien dengan BMD rendah.

Kemudian akan digunakan juga metode *Decision Tree* untuk melihat hubungan antara variabel prediktor dan variabel respon, serta dalam melakukan klasifikasi osteoporosis. Metode ini digunakan karena merupakan metode nonparametrik yang dapat

memilih variabel prediktor yang paling mempengaruhi variabel respon, serta metode ini lebih mudah untuk dipahami karena dapat disajikan secara visual dalam bentuk pohon. Menurut Gorunescu (2011) dalam bukunya yang berjudul *Data Mining: Concepts, Models, and Techniques*, metode *Decision Tree* memiliki beberapa kelebihan, diantaranya mudah untuk dipahami dan diinterpretasikan, bisa digunakan untuk data numerik dan kategorik, bersifat *robust*, proses cepat, bisa digunakan untuk dataset yang besar, dan berbagai kelebihan lainnya. Terdapat beberapa jenis metode *Decision Tree*, diantaranya adalah ID3 (*Iterative Dichotomiser*), C4.5 dan CART (*Classification and Regression Tree*). Perbedaan ketiga metode tersebut terdapat pada pengukuran pemisah variabelnya, dimana ID3 menggunakan ukuran *Information Gain*, C4.5 menggunakan *Gini Ratio*, sedangkan CART menggunakan *Gini Index* (Han, Kamber, & Pei, 2012). Jenis metode *Decision Tree* yang digunakan pada penelitian ini adalah metode CART yang menggunakan ukuran pemisah *Gini Index*, karena variabel prediktor yang digunakan merupakan variabel yang dipisahkan menjadi dua kelas (*binary split*) untuk menentukan apakah pasien memiliki BMD yang rendah atau normal.

Berdasarkan beberapa penjelasan tersebut, maka pada penelitian ini akan digunakan metode Regresi Logistik Biner, CTree dan SVM dalam mengklasifikasikan data hasil pengolahan DPR sebagai upaya deteksi dini pasien yang terkena osteoporosis, serta melihat bagaimana hubungan antara lebar *cortical bone* dan kekuatan garis pada *trabecular bone* dalam klasifikasi wanita *postmenopausal* dengan BMD rendah. Diharapkan penelitian ini dapat membantu instansi kesehatan dalam meningkatkan akurasi atau ketepatan dalam mendeteksi pasien yang terkena osteoporosis.

1.2 Rumusan Masalah

Deteksi dini osteoporosis perlu dilakukan agar dampak buruk yang bisa ditimbulkan dapat dihindari, khususnya pada wanita *postmenopausal* karena memiliki resiko terkena osteoporosis lebih besar daripada laki-laki. Namun ketersediaan alat untuk mendeteksi osteoporosis masih sangat sedikit dan harganya sangat mahal.

Disisi lain, banyak masyarakat, khususnya wanita *postmenopausal*, yang sering melakukan pemeriksaan gigi ke dokter gigi, sehingga terdapat banyak DPR yang dihasilkan. Berdasarkan penelitian yang telah dilakukan sebelumnya, beberapa indikator yang dapat digunakan dalam deteksi dini osteoporosis melalui DPR adalah lebar *cortical bone* dan kekuatan garis pada *trabecular bone*. Oleh karena itu, permasalahan utama yang akan dibahas dalam penelitian ini adalah metode apa yang terbaik antara metode Regresi Logistik Biner, CTree dan SVM dalam melakukan klasifikasi pasien dengan BMD rendah? Metode terbaik tersebut kemudian akan digunakan dalam program atau *graphic user interface* (GUI) untuk melakukan deteksi dini wanita *postmenopausal* dengan BMD rendah.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan utama yang ingin dicapai dalam penelitian ini adalah untuk mendapatkan metode terbaik antara metode Regresi Logistik Biner, *Classification Tree* (CTree) dan *Support Vector Machine* (SVM) dalam mengklasifikasikan wanita *postmenopausal* yang memiliki BMD rendah melalui DPR. Setelah didapat metode terbaik, maka akan dibuat juga program yang dapat mendeteksi wanita *postmenopausal* dengan BMD rendah dari DPR.

1.4 Manfaat

Manfaat yang diharapkan dari penelitian ini adalah dapat memberikan pengetahuan tambahan mengenai performa metode Regresi Logistik Biner, *Classification Tree* (CTree) dan *Support Vector Machine* (SVM) dalam melakukan klasifikasi pada salah satu data medis. Manfaat lain yang diharapkan dari penelitian ini adalah dapat memberikan program deteksi dini osteoporosis yang dapat digunakan di berbagai rumah sakit yang diharapkan mampu membantu proses diagnosis pasien yang terkena osteoporosis.

1.5 Batasan Masalah

Beberapa batasan yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Kategori yang digunakan dalam klasifikasi adalah dua kategori, yaitu 'BMD Normal' dan 'BMD Rendah', dimana jika seorang wanita *postmenopausal* terdeteksi 'Osteopenia' atau 'Osteoporosis', maka wanita tersebut akan dikategorikan memiliki 'BMD Rendah'.
2. Citra yang digunakan dalam analisis merupakan citra dengan format *.tif dengan resolusi citra 300 dpi (*dot per inch*).

(Halaman ini sengaja dikosongkan)

BAB II

TINJAUAN PUSTAKA

2.1 Osteoporosis

Osteoporosis adalah berkurangnya massa tulang dan perubahan arsitekturnya sampai batas patah tanpa keluhan. Faktor risiko wanita terkena osteoporosis lebih besar dibanding pria. Perkiraan global saat ini satu dari tiga wanita terkena keropos tulang. Sedangkan untuk pria, kira-kira satu banding empat. Penyakit ini tidak menimbulkan keluhan apa pun sampai tulang penderita patah atau retak (Pangkalan Ide, 2012).

Deteksi osteoporosis dilakukan dengan cara melakukan pengecekan kepadatan tulang atau *bone mineral density* (BMD). Salah satu alat yang dapat digunakan untuk mengukur BMD dan memprediksi risiko osteoporosis adalah menggunakan *Dual-energy X-ray Absorptiometry* (DXA) (Arifin et al., 2010).

2.2 Dental Panoramic Radiograph

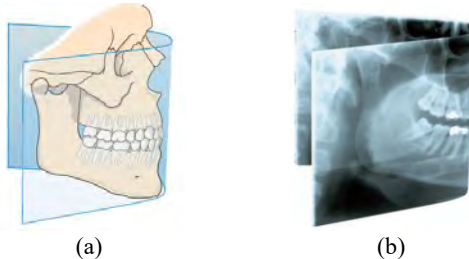
Radiograph adalah suatu citra yang diproduksi pada suatu lapisan tipis yang sensitif terhadap radiasi (*radiation sensitive film*), piringan fosfor (*phosphorous plate*), emulsi atau sensor digital dengan cara memancarkan radiasi ion (*American Dental Association*, 2012). Pada bidang kesehatan, *radiograph* dibuat menggunakan radiasi *electromagnetic X-rays* untuk melihat isi dalam tubuh manusia, seperti tulang, otak, *dental*, dan *fetus*.

Panoramic radiograph adalah proyeksi *extraoral* manusia, dimana seluruh citra rahang atas, rahang bawah, gigi, dan berbagai benda disekitarnya ditampilkan pada suatu citra, seolah-olah rahang menjadi rata pada suatu citra (*American Dental Association*, 2012). Ilustrasi posisi dalam pengambilan *panoramic radiograph* dapat dilihat pada Gambar 2.1 (a), sedangkan Gambar 2.1 (b) adalah hasil *panoramic radiograph* yang didapat.

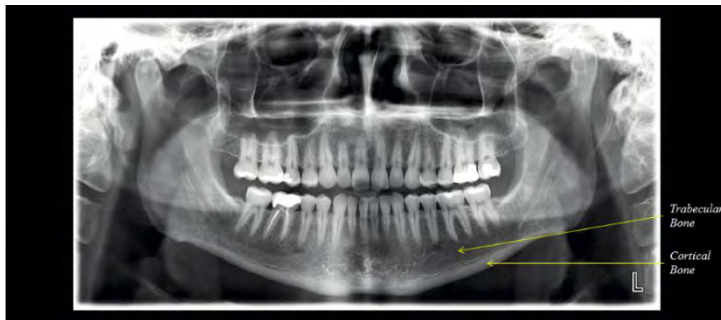
2.3 Struktur Tulang pada Rahang

Terdapat dua jenis tulang, yaitu *trabecular bone* (bersifat lunak, seperti spons) dan *cortical bone* (bersifat keras). *Trabecular bone* memberikan bantuan tenaga kepada tulang lain yang

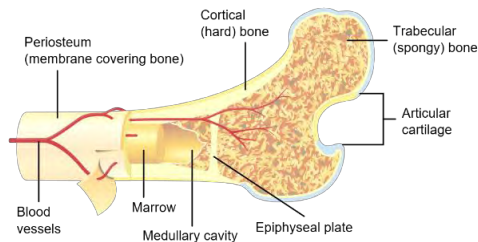
menyangga berat tubuh, sedangkan *cortical bone* berada di bagian terluar tulang dan memiliki bentuk seperti batang atau tangkai (Lian, Gorski, & Ott, 2004). Letak *trabecular bone* dan *cortical bone* pada tulang rahang dapat dilihat pada Gambar 2.2, sedangkan struktur kedua tulang tersebut lebih jelasnya dapat dilihat pada Gambar 2.3.



Gambar 2.1 (a) Posisi pengambilan *panoramic radiograph*, (b) hasil *panoramic radiograph*.



Gambar 2.2 Letak *trabecular bone* dan *cortical bone* pada rahang gigi.



Gambar 2.3 Struktur *trabecular bone* dan *cortical bone* pada tulang.

2.4 Image Processing

Image processing atau pengolahan citra adalah suatu cara untuk memproses atau memanipulasi citra menjadi citra lain yang kualitasnya lebih baik, agar citra yang mengalami gangguan mudah diinterpretasi, baik oleh manusia maupun mesin (Gonzalez, Woods, & Eddins, 2009). Tidak semua citra memiliki kualitas yang baik, terkadang terdapat beberapa citra yang mengalami penurunan kualitas, seperti mengandung cacat atau derau (*noise*), warna yang terlalu kontras, kurang tajam, kabur (*blur*), dan berbagai penurunan kualitas lainnya. Oleh karena itu, perlu adanya suatu *image processing* agar informasi yang didapat dari citra tersebut dapat dengan mudah diinterpretasikan.

Terdapat banyak operasi yang bisa dilakukan dalam melakukan *image processing*. Secara umum, operasi pengolahan citra dapat diklasifikasikan menjadi enam jenis operasi berikut.

1. Perbaikan kualitas citra (*image enhancement*)
 Operasi ini bertujuan untuk memperbaiki kualitas citra dengan cara memanipulasi parameter-parameter citra, sehingga ciri-ciri khusus yang terdapat dalam citra lebih ditonjolkan. Beberapa contoh proses perbaikan kualitas citra adalah sebagai berikut.
 - Perbaikan kontras gelap/terang.
 - Perbaikan tepian objek (*edge enhancement*).
 - Penajaman (*sharpening*).
 - Pemberian warna semu (*pseudocoloring*).
 - Penapisan derau (*noise filtering*).
2. Pemugaran citra (*image restoration*)
 Tujuan dari *image restoration* adalah untuk menghilangkan atau meminimumkan cacat pada citra. Berbeda dari *image enhancement*, pada operasi ini penyebab penurunan kualitas citra diketahui.
3. Pemampatan citra (*image compression*)
 Operasi ini dilakukan agar citra dapat direpresentasikan dalam bentuk yang lebih kompak, sehingga memori yang digunakan lebih sedikit, namun tetap menjaga kualitas dari citra.

4. Segmentasi citra (*image segmentation*)
Image segmentation bertujuan untuk memecah suatu citra ke dalam beberapa segmen dengan suatu kriteria tertentu.
5. Analisa citra (*image analysis*)
 Operasi ini bertujuan untuk menghitung besaran kuantitatif dari citra untuk menghasilkan deskripsinya. Proses ini dilakukan dengan cara mengekstraksi ciri-ciri tertentu yang membantu dalam mengidentifikasi objek. Proses segmentasi (*image segmentation*) kadang dibutuhkan untuk melokalisasi objek yang diinginkan dari objek sekelilingnya. Berikut adalah beberapa jenis *image analysis*.
 - Pendeteksian tepi objek (*edge detection*)
 - Ekstraksi batas (*boundary*)
 - Representasi daerah (*region*)
6. Rekonstruksi citra (*image reconstruction*)
Image reconstruction bertujuan untuk membentuk ulang objek dari beberapa citra hasil proyeksi. Operasi ini banyak digunakan dalam bidang medis, misalnya pembentukan citra ulang seluruh organ tubuh dari beberapa foto rontgen.

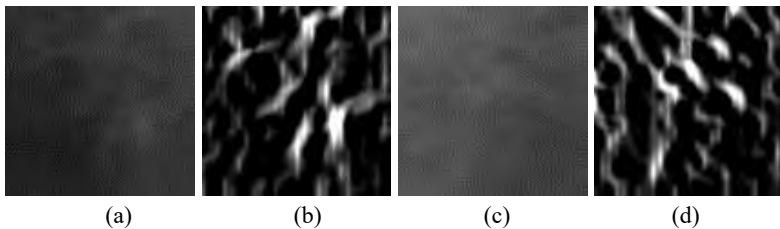
2.5 Line Strength Detection

Line Strength Detection merupakan salah satu metode pada *image processing* yang digunakan untuk mendeteksi struktur linier dari *trabecular bone* yang memiliki orientasi atau arah yang sama dengan akar gigi. Orientasi yang digunakan untuk rahang sebelah kiri adalah 0, 15, dan 30 derajat, sedangkan untuk rahang sebelah kanan adalah 330, 345, dan 360 derajat. Derajat ini nanti akan digunakan pada *Multiscale Line Operator* sebagai parameter θ .

Multiscale Line Operator adalah salah satu algoritma *Line Detection* yang digunakan untuk mendeteksi struktur linier atau garis lurus pada citra *mammographic*, yang digunakan bersama dengan metode *Line Detection* lainnya. Algoritma *Line Operator* ini terbukti memberikan hasil yang baik untuk kasus pemisahan sinyal dari citra yang memiliki *noise*, akurasi lebar struktur linier, dan lokalisasi (Arifin et al., 2010). Pada awal penggunaanya, algoritma *Multiscale Line Operator* digunakan untuk mendeteksi

serat asbes. Algoritma ini juga pernah digunakan untuk mendeteksi struktur linier pembuluh darah pada iris mata.

Pada penerapannya, *Multiscale Line Operator* dibantu oleh metode *Gaussian Pyramid* untuk menangani masalah ketika terdapat beberapa struktur garis yang memiliki lebar yang berbeda. Setelah proses *Multiscale Line Operator* dan *Gaussian Pyramid* dilakukan, citra hasil kedua proses tersebut kemudian akan diubah menjadi citra biner (citra hitam-putih), dimana *background* pada citra akan diubah menjadi hitam dengan nilai 0, dan objek pada citra akan diubah menjadi putih dengan nilai 1. Nilai kekuatan garis linier pada citra didapat dari citra biner dengan cara menghitung total pixel yang berwarna putih. Ilustrasi dari citra ROI dan citra hasil proses dapat dilihat pada Gambar 2.4 berikut.



Gambar 2.4 Ilustrasi citra ROI untuk rahang sebelah (a) kanan, (c) kiri, dan citra hasil pengolahan untuk rahang sebelah (b) kanan, (d) kiri.

2.6 Cortical Bone Width Measurement

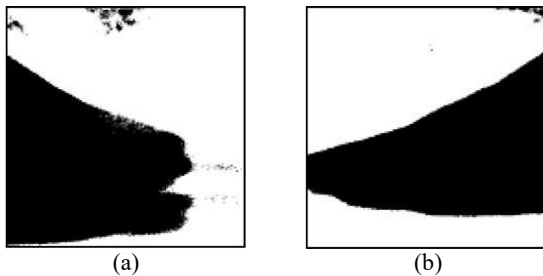
Pengukuran lebar terdiri dari tiga tahapan utama, yaitu *Image Enchantment*, *Cortical Margin Identification*, dan *Distance Measurement* (Arifin et al., 2006). Tahapan pertama yang dilakukan adalah menghilangkan area yang merupakan *background* menggunakan prinsip *Thresholding*, yaitu mengelompokkan suatu pixel menjadi suatu kelas jika memiliki nilai *gray level* lebih besar dari nilai *threshold* yang telah ditentukan, dan mengelompokkan pixel lainnya menjadi suatu kelompok lain jika memiliki nilai *gray level* lebih kecil dari nilai *threshold*. Pixel yang memiliki nilai *gray scale* lebih kecil dari 10% (yang menandakan efek cahaya) atau lebih besar dari 90% (yang menandakan gigi atau label teks pada citra) dari histogram *gray scale* keseluruhan pixel pada citra akan diabaikan, karena merupakan *outlier*. Pixel *outlier* tersebut kemu-

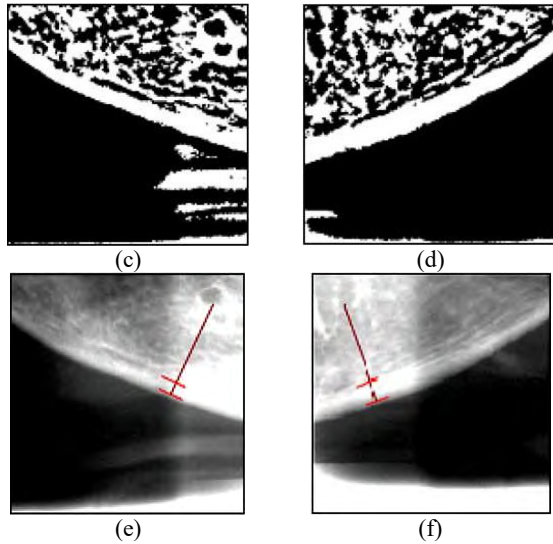
dian akan diubah menjadi bernilai 0, jika berada dibawah nilai persentil ke-10, atau 255, jika berada diatas persentil ke-90. Nilai *threshold* yang digunakan merupakan nilai yang berdasar pada varian antar kelas dan varian pada masing-masing kelas.

Citra hasil yang didapat dari algoritma *Thresholding* ini merupakan citra biner yang berisi objek penelitian yang akan dianalisis. Kemudian pada citra biner tersebut akan dilakukan proses *low-pass filtering* untuk mendapatkan citra yang hanya memiliki pixel dengan frekuensi rendah, dan mengurangi citra biner tersebut dengan citra hasil *low-pass filtering* sehingga didapat citra yang hanya berisi pixel dengan frekuensi yang tinggi. Citra yang dihasilkan tersebut kemudian dilakukan proses *high-pass filtering* untuk memperjelas batasan pada *cortical bone*. Proses terakhir pada tahapan pertama ini adalah melakukan *Morphological Open and Closed Operation* yang bertujuan untuk menggabungkan titik-titik kecil yang tersebar pada citra.

Tahapan kedua yang dilakukan adalah mengidentifikasi batasan luar pada citra yang telah dihasilkan pada tahap *Image Enhancement*. Kemudian pada setiap batasan akan dihitung *gradient* pada setiap titik hingga didapat *gradient* yang tepat untuk mengukur lebar *cortical bone* antar kedua batasan tersebut.

Tahapan terakhir adalah menghitung lebar *cortical bone* antar batasan yang telah ditentukan pada tahap *Cortical Margin Identification* menggunakan rumus jarak *Euclidean*. Jarak yang didapat ini, dalam pixel, kemudian ditransformasi menjadi jarak dalam milimeter. Ilustrasi tahapan yang dilakukan untuk menghitung lebar *cortical bone* dapat dilihat pada Gambar 2.5.





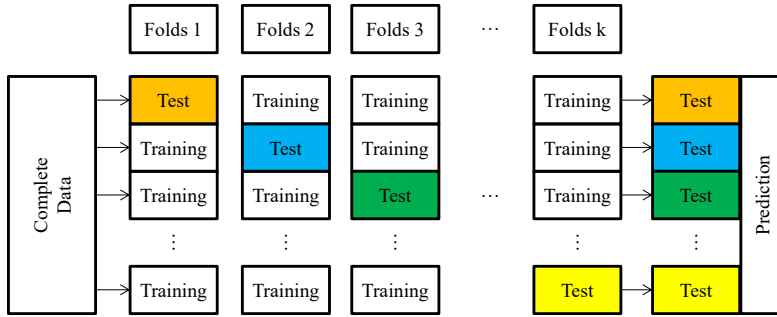
Gambar 2.5 Ilustrasi proses pengukuran lebar *cortical bone*, dimana (a), (b) citra hasil *Thresholding*, (c), (d) citra hasil *High-pass Filtering*, dan (e), (f) citra hasil pengukuran akhir pada rahang sebelah kanan dan rahang sebelah kiri.

2.7 K-folds Cross Validation (KCV)

Metode *K-folds Cross Validation* (KCV) merupakan suatu metode yang dapat diandalkan (*reliable*) untuk memprediksi kesalahan dalam suatu klasifikasi. Metode ini banyak digunakan oleh peneliti untuk mengurangi bias yang terjadi karena pengambilan sampel data yang akan digunakan. KCV secara berulang-ulang membagi data menjadi data *training* dan data *testing*, dimana setiap data berkesempatan menjadi data *testing* (Gokgoz & Subasi, 2015). Ilustrasi pembagian data menggunakan metode KCV dapat dilihat pada Gambar 2.6. Akurasi atau ketepatan klasifikasi yang dihasilkan dengan menggunakan metode KCV adalah sebagai berikut

$$CVA = \frac{1}{K} \sum_{k=1}^K A_k \quad (2.1)$$

dimana K adalah banyaknya *fold* yang digunakan, dan A_k adalah hasil akurasi yang dihasilkan pada setiap *fold*.



Gambar 2.6 Ilustrasi pembagian data menggunakan KCV.

Performa dari suatu metode klasifikasi dapat diukur dengan menggunakan *accuracy*, *sensitivity*, dan *specificity*. *Accuracy* adalah banyaknya pengamatan yang tepat diklasifikasikan sesuai kategorinya, *sensitivity* adalah banyaknya pengamatan yang berkategori positif yang tepat diklasifikasikan, dan *specificity* adalah banyaknya pengamatan yang berkategori negatif yang tepat diklasifikasikan. Perhitungan *accuracy*, *sensitivity*, dan *specificity* dapat dilakukan dengan menggunakan rumus sebagai berikut

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (2.2)$$

$$Sensitivity = \frac{TP}{TP+FP} \times 100\% \quad (2.3)$$

$$Specificity = \frac{TN}{TN+FN} \times 100\% \quad (2.4)$$

dimana *TP* adalah banyaknya pengamatan yang berkategori positif yang tepat diklasifikasikan ke kategori positif, *FP* adalah banyaknya pengamatan yang memiliki kategori positif yang diklasifikasikan ke kategori negatif, *TN* adalah banyaknya pengamatan yang berkategori negatif yang tepat diklasifikasikan ke kategori negatif, dan *FN* adalah banyaknya pengamatan yang berkategori negatif yang diklasifikasikan ke kategori positif.

Tahapan yang dilakukan dalam menggunakan KCV adalah sebagai berikut.

1. Tentukan nilai K atau banyak *fold*.
Langkah pertama adalah menentukan nilai K atau banyak *fold* yang akan digunakan. Nilai K yang umum digunakan adalah 5 atau 10.
2. Melakukan pengambilan data untuk data *training*.
Selanjutnya akan dilakukan pengambilan data secara random sebanyak $\frac{K-1}{K}n$, dimana n adalah banyak data, data yang diambil ini akan disebut data *training*.
3. Membuat model dengan menggunakan data *training*.
Kemudian akan dibuat model sesuai dengan metode yang digunakan, dimana data yang digunakan untuk membangun model adalah data *training* yang telah didapat sebelumnya.
4. Menghitung performa model.
Pada tahapan ini akan dihitung performa dari model yang didapat, baik pada data *training* maupun data *testing*. Untuk metode klasifikasi, performa model dapat dilihat menggunakan *accuracy*, *sensitivity*, dan *specificity*.
5. Ulangi langkah 2 hingga langkah 4 sebanyak K .
Mengulangi langkah 2 hingga langkah 4 sebanyak K -kali, sehingga semua data yang digunakan berkesempatan menjadi data *testing*.
6. Hitung rata-rata performa metode.
Setelah didapat sebanyak K -performa untuk data *training* dan K -performa untuk data *testing*, langkah terakhir adalah menghitung akurasi dari metode yang digunakan secara keseluruhan (CVA) dengan menggunakan persamaan (2.1).

2.8 Regresi Logistik Biner

Analisis regresi merupakan suatu metode yang digunakan untuk menjelaskan hubungan antara variabel respon (Y) dan variabel prediktor (X) (Draper & Smith, 1992). Jika variabel respon yang digunakan berupa data kategorik, maka analisis regresi yang digunakan adalah regresi logistik (Hosmer & Lemeshow, 2000). Sehingga analisis regresi logistik biner merupakan suatu metode yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor, dimana variabel respon yang digu-

nakan berupa data kategorik dengan banyak kategori adalah 2, yaitu ‘sukses’ ($Y = 1$) atau ‘gagal’ ($Y = 0$). Persamaan umum yang digunakan dalam regresi logistik biner adalah sebagai berikut

$$\ln \left(\frac{\pi(X)}{1-\pi(X)} \right) = \mathbf{X}\boldsymbol{\beta} \quad (2.5)$$

dengan $\pi(X)$ adalah peluang kejadian ‘sukses’, $\mathbf{X} = [\mathbf{1} \ X_1 \ X_2 \ \dots \ X_p]$, $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_p]$, dan p adalah banyak variabel prediktor yang digunakan. Berdasarkan persamaan (2.6), dapat dibentuk suatu persamaan untuk mencari $\pi(X)$ sebagai berikut

$$\pi(X) = \frac{\exp(\mathbf{X}\boldsymbol{\beta})}{1 + \exp(\mathbf{X}\boldsymbol{\beta})} \quad (2.6)$$

2.8.1 Pengecekan Asumsi Multikolinieritas

Istilah multikolinearitas pertama kali ditemukan oleh Ragnar Frisch, yang berarti adanya hubungan linear yang sempurna atau pasti di antara beberapa atau semua variabel penjelas (bebas) dari model regresi berganda. Selanjutnya istilah multikolinieritas digunakan dalam arti yang lebih luas, yaitu untuk terjadinya korelasi linear yang tinggi di antara variabel-variabel penjelas (prediktor) (X_1, X_2, \dots, X_p) (Setiawan & Kusri, 2010). Asumsi ini merupakan satu-satunya asumsi yang harus terpenuhi dalam menggunakan metode regresi logistik. Pengecekan asumsi multikolinieritas dapat dilakukan dengan melihat nilai *Variance Inflation Factor* (VIF) dari setiap variabel prediktor (X).

Nilai VIF didapat dengan cara melakukan regresi masing-masing variabel prediktor dengan variabel prediktor lainnya dan melihat nilai R^2 . Berikut adalah rumus untuk mencari nilai VIF dari variabel prediktor ke- j (X_j).

$$VIF = \frac{1}{1 - R^2(X_j)}, j = 1, 2, \dots, p \quad (2.7)$$

Pada data terjadi kasus multikolinieritas jika terdapat salah satu nilai VIF yang lebih besar dari 10.

2.8.2 Estimasi Parameter

Untuk mendapatkan model regresi logistik biner, maka perlu dilakukan proses estimasi parameter yang digunakan dalam model.

Estimasi akan dilakukan dengan menggunakan metode *Maximum Likelihood*, yaitu mengestimasi parameter β dengan cara memaksimumkan fungsi *likelihood*. Diketahui bahwa variabel respon (Y) yang digunakan berupa data kategorik dengan nilai 0 atau 1, sehingga dapat diketahui bahwa variabel respon ini mengikuti distribusi bernoulli dengan fungsi padat peluang (*probability mass function*) sebagai berikut

$$f(Y = y) = \pi(\mathbf{X})^y (1 - \pi(\mathbf{X}))^{1-y}. \quad (2.8)$$

Fungsi *likelihood* yang didapat dengan menggunakan fungsi padat peluang yang telah diketahui pada persamaan (2.8) adalah

$$\begin{aligned} l(\beta) &= \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \pi(\mathbf{X}_i)^{y_i} (1 - \pi(\mathbf{X}_i))^{1-y_i} \\ &= \prod_{i=1}^n \pi(\mathbf{X}_i)^{y_i} (1 - \pi(\mathbf{X}_i))^{-y_i} (1 - \pi(\mathbf{X}_i)) \\ &= \prod_{i=1}^n \left(\frac{\pi(\mathbf{X}_i)}{1 - \pi(\mathbf{X}_i)} \right)^{y_i} (1 - \pi(\mathbf{X}_i)) \end{aligned} \quad (2.9)$$

Fungsi *likelihood* tersebut kemudian diubah menjadi bentuk logaritma natural (\ln), sehingga didapat

$$\begin{aligned} \ln(l(\beta)) &= \ln \left(\prod_{i=1}^n \left(\frac{\pi(\mathbf{X}_i)}{1 - \pi(\mathbf{X}_i)} \right)^{y_i} (1 - \pi(\mathbf{X}_i)) \right) \\ L(\beta) &= \sum_{i=1}^n \ln \left(\left(\frac{\pi(\mathbf{X}_i)}{1 - \pi(\mathbf{X}_i)} \right)^{y_i} (1 - \pi(\mathbf{X}_i)) \right) \\ &= \sum_{i=1}^n y_i \ln \left(\frac{\pi(\mathbf{X}_i)}{1 - \pi(\mathbf{X}_i)} \right) + \ln(1 - \pi(\mathbf{X}_i)) \\ &= \sum_{i=1}^n y_i \mathbf{X}_i \beta + \ln \left(1 - \frac{\exp(\mathbf{X}_i \beta)}{1 + \exp(\mathbf{X}_i \beta)} \right) \\ &= \sum_{i=1}^n y_i \mathbf{X}_i \beta + \ln \left(\frac{1}{1 + \exp(\mathbf{X}_i \beta)} \right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n y_i \mathbf{X}_i \boldsymbol{\beta} + \ln(1 + \exp(\mathbf{X}_i \boldsymbol{\beta}))^{-1} \\
&= \sum_{i=1}^n y_i \mathbf{X}_i \boldsymbol{\beta} - \ln(1 + \exp(\mathbf{X}_i \boldsymbol{\beta})) \quad (2.10)
\end{aligned}$$

dengan n adalah banyak pengamatan.

Tahapan selanjutnya adalah mendapatkan nilai $\boldsymbol{\beta}$ yang menyebabkan nilai dari fungsi *likelihood* bernilai ekstrim, dengan cara mencari turunan pertama $L(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$. Hasil turunan yang didapat adalah

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{g}(\boldsymbol{\beta}) &= \sum_{i=1}^n y_i \mathbf{X}_i - \frac{\exp(\mathbf{X}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta})} \mathbf{X}_i \\
&= \sum_{i=1}^n y_i \mathbf{X}_i - \pi(\mathbf{X}_i) \mathbf{X}_i = 0 \quad (2.11)
\end{aligned}$$

jika dilanjutkan ke turunan kedua, maka didapat

$$\begin{aligned}
\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}^T} \left(\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right) \\
&= \frac{\partial}{\partial \boldsymbol{\beta}^T} \left(\sum_{i=1}^n y_i \mathbf{X}_i - \frac{\exp(\mathbf{X}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta})} \mathbf{X}_i \right) \\
&= \sum_{i=1}^n 0 - \frac{\mathbf{X}_i^T \exp(\mathbf{X}_i \boldsymbol{\beta}) (1 + \exp(\mathbf{X}_i \boldsymbol{\beta}))}{(1 + \exp(\mathbf{X}_i \boldsymbol{\beta}))^2} \mathbf{X}_i \\
&\quad - \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}) \mathbf{X}_i^T \exp(\mathbf{X}_i \boldsymbol{\beta})}{(1 + \exp(\mathbf{X}_i \boldsymbol{\beta}))^2} \mathbf{X}_i \\
&= \sum_{i=1}^n - \frac{\mathbf{X}_i^T \exp(\mathbf{X}_i \boldsymbol{\beta})}{(1 + \exp(\mathbf{X}_i \boldsymbol{\beta}))^2} \mathbf{X}_i \\
&= - \sum_{i=1}^n \mathbf{X}_i^T \frac{\exp(\mathbf{X}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta})} \frac{1}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta})} \mathbf{X}_i
\end{aligned}$$

$$\begin{aligned}
&= - \sum_{i=1}^n \mathbf{X}_i^T \pi(\mathbf{X}_i)(1 - \pi(\mathbf{X}_i)) \mathbf{X}_i \\
\mathbf{H}(\boldsymbol{\beta}) &= - \sum_{i=1}^n \mathbf{X}_i^T \text{Var}(\mathbf{X}_i) \mathbf{X}_i
\end{aligned} \tag{2.12}$$

dimana $\mathbf{H}(\boldsymbol{\beta})$ adalah matriks Hessian. Karena turunan kedua selalu bernilai negatif, maka didapat bahwa nilai $\boldsymbol{\beta}$ membuat fungsi *likelihood* bernilai ekstrim maksimum. Namun karena hasil turunan pertama pada persamaan (2.11) tidak mendapatkan hasil yang eksplisit atau rumus untuk mencari nilai $\boldsymbol{\beta}$ tidak didapat, maka akan digunakan metode numerik, yaitu metode iterasi *Newton Rhapson* untuk mencari estimasi dari parameter $\boldsymbol{\beta}$. Langkah-langkah yang dilakukan dalam iterasi *Newton Rhapson* adalah sebagai berikut.

1. Menentukan nilai inisiasi awal untuk $\boldsymbol{\beta}$

Inisiasi awal untuk parameter $\boldsymbol{\beta}$, yaitu $\boldsymbol{\beta}^{(0)}$, dapat dilakukan dengan menggunakan rumus OLS (*Ordinary Leas Square*) pada persamaan regresi linier berganda. Rumus yang digunakan adalah sebagai berikut

$$\boldsymbol{\beta}^{(0)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \tag{2.13}$$

2. Melakukan iterasi *Newton Rhapson*

Pada tahapan ini akan dilakukan iterasi dengan menggunakan nilai $\boldsymbol{\beta}^{(l)}$ yang didapat, dimana $\boldsymbol{\beta}^{(l)}$ adalah nilai estimasi parameter $\boldsymbol{\beta}$ pada iterasi ke- l . Rumus yang digunakan dalam iterasi adalah sebagai berikut

$$\boldsymbol{\beta}^{(l+1)} = \boldsymbol{\beta}^{(l)} + (-\mathbf{H}(\boldsymbol{\beta}^{(l)}))^{-1} \mathbf{g}(\boldsymbol{\beta}^{(l)}) \tag{2.14}$$

dimana $\mathbf{H}(\boldsymbol{\beta}^{(l)})$ adalah matriks Hessian yang telah didapat pada persamaan (2.14) dan $\mathbf{g}(\boldsymbol{\beta}^{(l)})$ adalah vektor hasil turunan pertama yang telah didapat pada persamaan (2.11).

3. Cek apakah estimasi parameter yang didapat telah konvergen
Setelah didapat estimasi parameter baru ($\boldsymbol{\beta}^{(l+1)}$), langkah selanjutnya adalah melakukan pengecekan apakah estimasi parameter yang didapat telah konvergen atau tidak. Apabila telah didapat estimasi parameter yang konvergen, maka iterasi

berhenti. Pengecekan dilakukan dengan menggunakan rumus berikut

$$\|\beta^{(l+1)} - \beta^{(l)}\| \leq \varepsilon \quad (2.15)$$

dimana ε adalah suatu bilangan yang sangat kecil, yang ditentukan oleh peneliti.

2.8.3 Uji Signifikansi Parameter

Setelah didapat nilai estimasi parameter yang digunakan dalam model, selanjutnya akan dilakukan pengujian apakah parameter yang didapat tersebut signifikan atau tidak dalam menjelaskan hubungan antara variabel prediktor dengan variabel respon. Terdapat dua jenis pengujian yang akan dilakukan, yaitu pengujian secara serentak dan secara parsial. Hipotesis yang digunakan dalam uji serentak adalah sebagai berikut,

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1: \text{Minimal ada satu } \beta_j \neq 0, j = 1, 2, \dots, p.$$

statistik uji yang digunakan adalah sebagai berikut

$$G = -2 \ln \left(\frac{\left(\frac{n_0}{n}\right)^{n_0} \left(\frac{n_1}{n}\right)^{n_1}}{\prod_{i=1}^n (\hat{\pi}(X_i))^{y_i} (1 - \hat{\pi}(X_i))^{1-y_i}} \right) \quad (2.16)$$

dengan n_0 adalah banyaknya pengamatan yang bernilai $Y = 0$ dan n_1 adalah banyaknya pengamatan yang bernilai $Y = 1$. Hipotesis H_0 ditolak jika $G > \chi_{\frac{\alpha}{2}, p}^2$ atau $p - \text{value} < \alpha$.

Jika pada pengujian serentak menghasilkan kesimpulan tolak H_0 , maka pengujian akan dilanjutkan dengan uji parsial, dimana hipotesis yang digunakan adalah sebagai berikut,

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0, j = 1, 2, \dots, p.$$

statistik uji yang digunakan adalah,

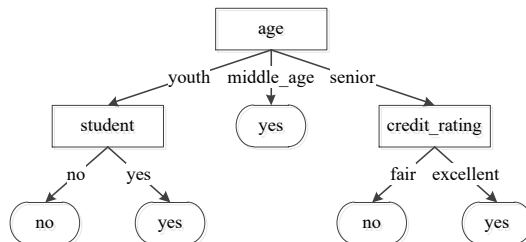
$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (2.17)$$

Hipotesis H_0 ditolak jika $|Z| > Z_{\frac{\alpha}{2}}$ atau $p - \text{value} < \alpha$.

2.9 Classification Tree (CTree)

Algoritma *Decision Tree* pertama kali dikembangkan oleh J. Ross Quinlan, peneliti bidang *machine learning*, pada awal tahun 1980, yang mengembangkan jenis *Decision Tree* ID3 (*Iterative Dichotomiser*). Metode ini kemudian dikembangkan menjadi C4.5 (pengembangan dari ID3) yang kemudian dasar untuk pengembangan algoritma pembelajaran lanjutan. Kemudian pada tahun 1984, L. Breiman, J. Friedman, R. Olshen, dan C. Stone menerbitkan buku *Classification and Regression Tree* (CART) yang merupakan salah satu jenis *Decision Tree* dengan kelas biner (*Binary Decision Tree*). ID3 dan CART ditemukan secara independen pada waktu yang hampir bersamaan, namun memiliki cara kerja yang sama. Perbedaan antara ID3, C4.5, dan CART terletak pada jenis pengukuran yang digunakan untuk memilih variabel yang akan digunakan pada pohon, dimana ID3 menggunakan ukuran *Information Gain*, C4.5 menggunakan ukuran *Gini Ratio*, sedangkan CART menggunakan *Gini Index*.

Decision Tree adalah suatu bentuk pohon yang menyerupai diagram alir, dimana *internal node* menunjukkan variabel prediktor yang digunakan sebagai pemisah yang dihubungkan oleh cabang (*branch*), dan setiap *leaf node* merupakan kelas hasil klasifikasi (Han et al., 2012). Salah satu contoh *Decision Tree* untuk kasus data pembelian komputer dapat dilihat pada Gambar 2.7 berikut.



Gambar 2.7 Contoh pohon klasifikasi pada kasus pembelian komputer.

CART merupakan salah satu jenis *Decision Tree* dimana banyak kelas yang digunakan dalam analisis adalah dua (*Binary Decision Tree*). Ukuran pemilihan variabel yang digunakan dalam

CART adalah *Gini Index*, dimana rumus yang digunakan adalah sebagai berikut.

$$Gini(D) = 1 - \sum_{h=1}^H p_{D_h}^2 \quad (2.18)$$

dengan p_{D_h} adalah proporsi partisi data D pada kelas C_h (Han et al., 2012). Setelah didapat nilai *Gini Index*, selanjutnya akan dicari nilai *goodness of split* untuk menentukan variabel mana yang akan digunakan terlebih dahulu pada simpul. Rumus yang digunakan untuk mendapatkan *goodness of split* adalah sebagai berikut

$$\Delta Gini(X) = Gini(D) - p_{D_{Left}} Gini(D_{Left}) - p_{D_{Right}} Gini(D_{Right}) \quad (2.19)$$

dengan

- $\Delta Gini(X)$: *Goodness of split* dari variabel X,
- $Gini(D)$: *Gini Index* pada partisi data D,
- $p_{D_{Left}}$: Proporsi pengamatan pada simpul kiri,
- $p_{D_{Right}}$: Proporsi pengamatan pada simpul kanan,
- $Gini(D_{Left})$: *Gini Index* pada simpul kiri,
- $Gini(D_{Right})$: *Gini Index* pada simpul kanan.

Variabel yang dipilih sebagai pemilah adalah variabel yang memiliki nilai *goodness of split* terbesar, karena variabel tersebut mampu mereduksi heterogenitas lebih besar.

Jika variabel prediktor yang digunakan merupakan data kategorik, maka pemisah simpul sebelumnya menjadi simpul kanan dan simpul kiri dapat menggunakan kategori pada variabel tersebut. Namun jika variabel prediktor yang digunakan berskala rasio atau merupakan data numerik, maka digunakan berbagai kemungkinan nilai tengah (median) antar setiap data yang telah diurutkan sebagai pemisah simpul, kemudian akan dipilih nilai tengah yang menghasilkan nilai *goodness of fit* terbesar. Berikut adalah tahapan-tahapan yang dilakukan dalam membuat pohon klasifikasi.

1. Menghitung nilai *Gini Index* untuk variabel respon (Y)
Langkah pertama adalah menghitung nilai *Gini Index* dari variabel respon dengan menggunakan persamaan (2.18).

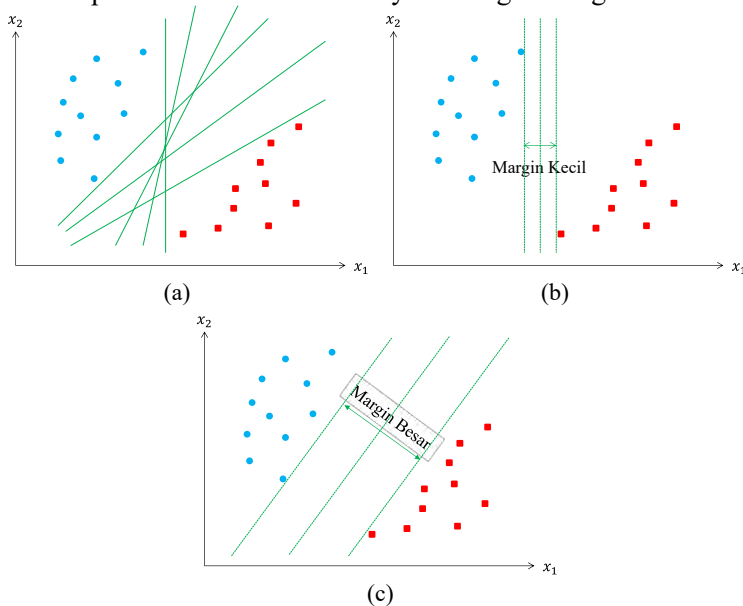
2. Menentukan nilai *threshold* untuk setiap variabel prediktor yang digunakan
Kemudian menentukan nilai *threshold* atau pemisah simpul dari masing-masing variabel prediktor yang digunakan, dimana jika variabel prediktor merupakan data kategorik, maka pemisah simpul adalah kategori pada variabel tersebut, dan jika variabel prediktor merupakan data numerik, digunakan nilai tengah antar data sebagai pemisah simpul.
3. Menghitung nilai *goodness of split* untuk setiap variabel prediktor
Setelah didapat pemisah simpul, maka akan dihitung nilai *goodness of split* menggunakan persamaan (2.19).
4. Memilih variabel prediktor yang digunakan sebagai simpul
Setelah didapat beberapa nilai *goodness of split* dari setiap variabel prediktor yang digunakan, maka akan dipilih variabel prediktor yang menghasilkan nilai *goodness of split* terbesar sebagai simpul.
5. Ulangi langkah 2 sampai langkah 4 hingga semua variabel prediktor digunakan atau data sudah terbagi secara sempurna
Tahapan 2 hingga tahapan 4 dilakukan kembali menggunakan variabel yang belum digunakan pada simpul. Tahapan ini dilakukan hingga semua variabel prediktor telah digunakan sebagai simpul pada pohon klasifikasi, atau pohon sudah dapat membagi data secara sempurna.

2.10 Support Vector Machine (SVM)

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vladimir Vapnik, Bernhard Boser, dan Isabelle Guyon pada tahun 1992, walaupun pengembangan metode ini sudah ada sejak sekitar tahun 1960. SVM merupakan suatu algoritma untuk mencari *hyperplane* yang dapat memisahkan data antara dua kelas oleh pemisah linier secara optimal. Meskipun waktu proses yang diperlukan cenderung lama, akurasi yang dihasilkan cenderung tinggi dan tidak terjadi kasus *overfitting* (Han et al., 2012).

2.10.1 Kasus Data yang Bisa Dipisah Secara Linier

Salah satu contoh kasus dasar pada klasifikasi adalah data yang bisa dipisahkan secara linier oleh suatu garis lurus. Ilustrasi data yang dapat dipisahkan oleh garis lurus dapat dilihat pada Gambar 2.8 (a), dimana terdapat banyak garis yang bisa digunakan untuk memisahkan data tersebut ke dalam kelas yang bersesuaian. Contoh garis pemisah yang dapat digunakan dapat dilihat pada Gambar 2.8 (b) dan Gambar 2.8 (c), dimana dapat diketahui bahwa pemisah pada Gambar 2.8 (c) lebih baik untuk digunakan karena memiliki margin yang lebih besar, sehingga data dapat secara akurat dipisahkan ke dalam kelasnya masing-masing.



Gambar 2.8 (a) Ilustrasi beberapa *hyperplane* yang bisa digunakan pada data yang bisa dipisah secara linier, dimana (b) *hyperplane* yang menghasilkan margin yang kecil, (c) *hyperplane* yang menghasilkan margin yang besar.

Seperti yang dapat dilihat pada Gambar 2.8 (a), terdapat sebanyak tak hingga garis pemisah yang bisa dibuat untuk memisahkan data kategori +1 (warna biru) dengan data kategori -1

(warna merah). Namun akan dicari suatu garis pemisah yang terbaik, yang bisa meminimumkan kesalahan klasifikasi, yaitu garis yang menghasilkan margin terbesar. Margin adalah jarak antara *hyperplane* atau pemisah dengan data terdekat dari masing-masing kelas data. Persamaan untuk *hyperplane* dapat ditulis sebagai berikut

$$\mathbf{W} \cdot \mathbf{X} + b = 0, \quad (2.20)$$

dengan $\mathbf{W} = \{w_1, w_2, \dots, w_p\}$ adalah vektor pembobot, p adalah banyak variabel X , dan b adalah suatu konstanta atau biasa disebut dengan bias. Persamaan (2.22) dapat dimodifikasi sehingga didapat persamaan untuk setiap sisi margin sebagai berikut

$$M_1 : \mathbf{W} \cdot \mathbf{X} + b \geq 1 \text{ untuk } y_m = +1, \quad (2.21)$$

$$M_2 : \mathbf{W} \cdot \mathbf{X} + b \leq -1 \text{ untuk } y_m = -1. \quad (2.22)$$

Data yang berada pada daerah M_1 akan dimasukkan ke kelas +1, sedangkan data yang berada pada daerah M_2 akan dimasukkan ke kelas -1. Jika persamaan tersebut dikalikan dengan dengan dengan masing-masing nilai kelasnya, y_m , maka didapat persamaan sebagai berikut

$$y_m(\mathbf{W} \cdot \mathbf{X} + b) \geq 1, \forall_m \quad (2.23)$$

Data yang berada tepat pada tepi margin, yaitu M_1 dan M_2 , disebut sebagai *support vectors*.

Berdasarkan persamaan (2.23) dapat diketahui bahwa jarak terdekat *hyperplane* ke tepi margin adalah $\frac{1}{\|\mathbf{W}\|}$, dimana $\|\mathbf{W}\|$ adalah *Euclidian norm* dari \mathbf{W} , dengan rumus $\sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$. Sehingga jarak antar tepi margin adalah $\frac{2}{\|\mathbf{W}\|}$. Karena akan dicari *hyperplane* yang memberikan margin maksimal, yaitu $\max \frac{2}{\|\mathbf{W}\|}$, hal ini kongruen dengan mencari nilai minimal dari $\min \frac{\|\mathbf{W}\|}{2}$, atau $\min \frac{\|\mathbf{W}\|^2}{2}$. Maka fungsi obyektif pada permasalahan ini adalah

$$\min \frac{\|\mathbf{W}\|^2}{2} \quad (2.24)$$

dengan fungsi batasan sebagai berikut.

$$\sum_{m=1}^M \alpha_m [y_m (\mathbf{W} \cdot \mathbf{X}_m + b) - 1] \quad (2.25)$$

Kemudian dengan menggunakan *Lagrange Multiplier*, didapat persamaan sebagai berikut

$$L_{pd} = \frac{||\mathbf{W}'||^2}{2} - \sum_{m=1}^M \alpha_m [y_m (\mathbf{W} \cdot \mathbf{X}_m + b) - 1]. \quad (2.26)$$

Selanjutnya dengan menggunakan kondisi *Karush-Kuhn-Tucker* (KKT), yaitu

$$\frac{\partial L_{pd}}{\partial \mathbf{W}} = 0 \leftrightarrow \mathbf{W} - \sum_{m=1}^M \alpha_m y_m \mathbf{X}_m = 0 \quad (2.27)$$

$$\mathbf{W} = \sum_{m=1}^M \alpha_m y_m \mathbf{X}_m$$

$$\frac{\partial L_{pd}}{\partial b} = 0 \leftrightarrow 0 - \sum_{m=1}^M \alpha_m y_m = 0 \leftrightarrow \sum_{m=1}^M \alpha_m y_m = 0 \quad (2.28)$$

$$\alpha_m [y_m (\mathbf{W} \cdot \mathbf{X}_m + b) - 1] = 0, \text{ dimana } \alpha_m \geq 0 \quad (2.29)$$

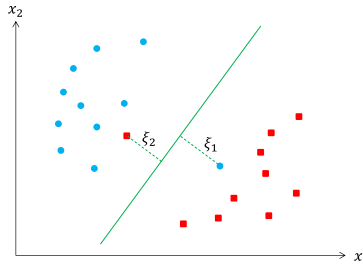
dengan mensubstitusi persamaan (2.27) hingga persamaan (2.29) ke persamaan (2.26), maka didapat persamaan sebagai berikut,

$$L_d = -\frac{1}{2} \sum_{m_1=1}^M \sum_{m_2=1}^M \alpha_{m_1} \alpha_{m_2} y_{m_1} y_{m_2} (\mathbf{X}_{m_1} \cdot \mathbf{X}_{m_2}) + \sum_{m=1}^M \alpha_m \quad (2.30)$$

kemudian dengan mensubstitusi nilai y_{m_1} , y_{m_2} , \mathbf{X}_{m_1} , dan \mathbf{X}_{m_2} ke persamaan (2.30), didapat suatu persamaan L_d yang kemudian akan digunakan untuk mendapatkan nilai-nilai α_m (*support vectors*) yang membuat L_d optimum dengan cara mencari turunan parsial L_d terhadap α .

2.10.2 Kasus Data yang Tidak Bisa Dipisah Secara Linier

Kasus data yang bisa dipisahkan secara linier merupakan kasus yang sulit untuk ditemui. Umumnya, terdapat beberapa kelas data yang berada pada daerah kelas data lainnya, kasus ini disebut *linearly non-separable data* atau data yang tidak bisa dipisah secara linier. Salah satu ilustrasi untuk kasus ini dapat dilihat pada Gambar 2.9 berikut.



Gambar 2.9 Ilustrasi kasus data yang tidak bisa dipisah secara linier.

Untuk kasus ini, maka pencarian *hyperplane* yang optimal akan memperhatikan data-data yang tidak berada pada kelasnya (*misclassification error*), yang dilambangkan dengan ξ . Sehingga persamaan (2.25) menjadi

$$y_m(\mathbf{W} \cdot \mathbf{X} + b) \geq 1 - \xi_m, \forall_m, \quad (2.31)$$

dan didapat persamaan *Lagrange Multiplier* sebagai berikut

$$\begin{aligned} L_{pd} = & \frac{\|\mathbf{W}\|^2}{2} + C \sum_m^M \xi_m \\ & - \sum_{m=1}^M \alpha_m [y_m(\mathbf{W} \cdot \mathbf{X}_m + b) - 1 + \xi_m] \\ & - \sum_{m=1}^M \beta_m \xi_m, \end{aligned} \quad (2.32)$$

dengan C adalah suatu nilai pengali *Lagrange*. Kemudian dengan menggunakan kondisi *Karush-Kuhn-Tucker* (KKT), yaitu

$$\frac{\partial L_{pd}}{\partial \mathbf{W}} = 0 \leftrightarrow \mathbf{W} - \sum_{m=1}^M \alpha_m y_m \mathbf{X}_m = 0 \quad (2.33)$$

$$\mathbf{W} = \sum_{m=1}^M \alpha_m y_m \mathbf{X}_m$$

$$\frac{\partial L_{pd}}{\partial b} = 0 \leftrightarrow 0 - \sum_{m=1}^M \alpha_m y_m = 0 \leftrightarrow \sum_{m=1}^M \alpha_m y_m = 0 \quad (2.34)$$

$$\frac{\partial L_{pd}}{\partial \xi_m} = 0 \leftrightarrow 0 + C - \alpha_m - \beta_m = 0 \leftrightarrow C = \alpha_m + \beta_m \quad (2.35)$$

dengan mensubstitusi persamaan (2.33) hingga persamaan (2.35) ke persamaan (2.32), didapat persamaan sebagai berikut,

$$L_d = -\frac{1}{2} \sum_{m_1=1}^M \sum_{m_2=1}^M \alpha_{m_1} \alpha_{m_2} y_{m_1} y_{m_2} (\mathbf{X}_{m_1} \cdot \mathbf{X}_{m_2}) + \sum_{m=1}^M \alpha_m \quad (2.36)$$

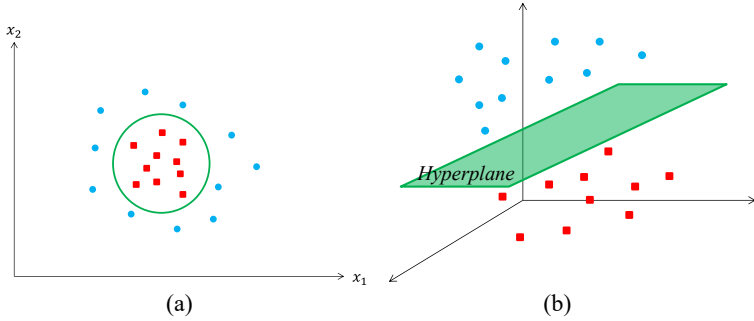
Selanjutnya dengan menggunakan langkah yang sama pada kasus data yang bisa terpisah secara linier, yaitu mensubstitusikan nilai y_{m_1} , y_{m_2} , \mathbf{X}_{m_1} , dan \mathbf{X}_{m_2} ke persamaan (2.36), maka didapat suatu persamaan L_d yang kemudian akan digunakan untuk mendapatkan nilai-nilai α_m (*support vectors*) yang membuat L_d optimum dengan cara mencari turunan parsial L_d terhadap α .

2.10.3 Fungsi Kernel pada SVM untuk Kasus Data Non-Linier

Pada kasus nyata, sangat jarang dijumpai data yang dapat terpisah secara linier, oleh karena itu digunakan suatu fungsi kernel (Φ) untuk memetakan data ke dalam ruang vektor yang berdimensi tinggi, seperti yang telah diilustrasikan pada Gambar 2.10, sehingga data hasil transformasi bisa terpisah secara linier. Setelah dilakukan transformasi, langkah selanjutnya adalah menemukan *support vector* dari data yang sudah di transformasi ke ruang yang berdimensi tinggi, yang bisa didapat menggunakan pengembangan dari persamaan (2.30) atau (2.36), yaitu

$$L_d = -\frac{1}{2} \sum_{m_1=1}^M \sum_{m_2=1}^M \alpha_{m_1} \alpha_{m_2} y_{m_1} y_{m_2} \left(\Phi(\mathbf{X}_{m_1}) \cdot \Phi(\mathbf{X}_{m_2}) \right) + \sum_{m=1}^M \alpha_m \quad (2.37)$$

dengan $\Phi(\mathbf{X}_{m_1})$ atau $\Phi(\mathbf{X}_{m_2})$ adalah data hasil transformasi.



Gambar 2.10 (a) Ilustrasi kasus data non-linier, (b) hasil transformasi data beserta *hyperplane* yang didapat.

Namun transformasi Φ tidak dapat diketahui dan sangat sulit dipahami, sehingga perhitungan *dot product* dapat secara implisit digantikan oleh fungsi kernel, sehingga didapat,

$$L_d = -\frac{1}{2} \sum_{m_1=1}^M \sum_{m_2=1}^M \alpha_{m_1} \alpha_{m_2} y_{m_1} y_{m_2} K(\mathbf{X}_{m_1}, \mathbf{X}_{m_2}) + \sum_{m=1}^M \alpha_m \quad (2.38)$$

dengan $K(\mathbf{X}_{m_1}, \mathbf{X}_{m_2})$ adalah fungsi kernel yang digunakan. Beberapa fungsi kernel yang akan digunakan pada penelitian ini adalah fungsi kernel *linier*, fungsi kernel *polynomial* dengan parameter *degree*, dan fungsi kernel *radial basis* dengan parameter *gamma*. Rincian rumus dan parameter yang digunakan pada setiap fungsi kernel dapat dilihat pada Tabel 2.1 berikut.

Tabel 2.1 Beberapa fungsi kernel yang umum digunakan pada SVM

Fungsi Kernel	Rumus $K(\mathbf{X}_{m_1}, \mathbf{X}_{m_2})$	Parameter
<i>Linear</i>	$1 + \mathbf{X}_{m_1} \cdot \mathbf{X}_{m_2}^T$	-
<i>Polynomial</i>	$(1 + \mathbf{X}_{m_1} \cdot \mathbf{X}_{m_2}^T)^d$	d
<i>Radial Basis</i>	$\exp\left(-\frac{(\mathbf{X}_{m_1} - \mathbf{X}_{m_2}) \cdot (\mathbf{X}_{m_1} - \mathbf{X}_{m_2})^T}{2\sigma^2}\right)$	σ

2.11 Penelitian Sebelumnya

Penelitian terkait deteksi osteoporosis telah banyak dilakukan sebelumnya. Deteksi osteoporosis melalui *cortical bone* telah dilakukan oleh Arifin et al. (2006) dan Arhandi, Arifin, & Khotimah (2014). Deteksi osteoporosis juga dilakukan melalui *trabecular bone*, seperti yang dilakukan oleh Abidin & Arifin (2008), Arifin et al. (2010), dan Wihandika (2013). Namun penelitian tersebut hanya mendeteksi osteoporosis berdasarkan satu indikator atau variabel saja dan deteksi osteoporosis yang dilakukan hanya berdasarkan suatu nilai *threshold*.

BAB III

METODOLOGI PENELITIAN

3.1 Sumber Data

Penelitian ini menggunakan data sekunder, dimana DPR yang digunakan diperoleh dari penelitian yang telah dilakukan oleh Arifin dkk (2006) yang berjudul '*Computer-aided System for Measuring the Mandibular Cortical Width on Dental Panoramic Radiographs in Identifying Postmenopausal Women with Low Bone Mineral Density*'. DPR yang digunakan merupakan citra yang diambil dari 100 wanita *postmenopausal* yang berusia lebih dari 50 tahun.

3.2 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini merupakan variabel yang berisi informasi yang bisa didapat dari hasil pengolahan DPR. Berikut adalah variabel yang digunakan pada penelitian ini.

Tabel 3.1 Variabel yang digunakan dalam penelitian

Variabel	Keterangan	Skala Data
Y	Kategori BMD wanita <i>postmenopausal</i> yang diteliti, dengan rincian 0 : BMD Rendah, sebanyak 46 pengamatan. 1 : BMD Normal, sebanyak 54 pengamatan.	Nominal
X ₁	Lebar <i>Cortical Bone (Right)</i> . Merupakan ukuran lebar <i>cortical bone</i> (dalam mm) pada bagian kanan rahang gigi.	Rasio
X ₂	Lebar <i>Cortical Bone (Left)</i> . Merupakan ukuran lebar <i>cortical bone</i> (dalam mm) pada bagian kiri rahang gigi.	Rasio
X ₃	Kekuatan garis pada <i>Trabecular Bone (Right)</i> . Yaitu kekuatan garis pada tulang trabecular (dalam pixel) pada bagian kanan rahang gigi.	Rasio
X ₄	Kekuatan garis pada <i>Trabecular Bone (Left)</i> . Yaitu kekuatan garis pada tulang trabecular (dalam pixel) pada bagian kiri rahang gigi.	Rasio

3.3 Langkah Analisis

Tahapan yang dilakukan dalam penelitian ini setelah didapat DPR adalah sebagai berikut.

1. Melakukan *preprocessing* DPR untuk mendapatkan data yang bisa diolah. Beberapa tahapan *preprocessing* yang dilakukan adalah sebagai berikut.
 - a. Menghitung kekuatan garis pada *trabecular bone*.
Tahapan *preprocessing* yang pertama kali dilakukan adalah menghitung kekuatan garis pada *trabecular bone*, baik pada rahang gigi sebelah kiri maupun sebelah kanan. Perhitungan ini dilakukan dengan menggunakan metode *Line Strength Detection*, dengan tahapan sebagai berikut.
 1. Melakukan algoritma *Multiscale Line Operator* untuk mendeteksi struktur linier dari *trabecular bone* yang memiliki orientasi yang sama dengan akar gigi.
 2. Menggunakan metode *Gaussian Pyramid* untuk menangani masalah ketika terdapat beberapa struktur garis yang memiliki lebar yang berbeda.
 3. Menghitung kekuatan garis pada *trabecular bone* dengan cara menghitung total pixel yang berwarna putih dari citra biner hasil *Multiscale Line Operator* dan *Gaussian Pyramid*.
 - b. Mengukur lebar *cortical bone*.
Pada tahapan ini, akan dilakukan pengukuran lebar *cortical bone*, baik pada rahang sebelah kiri maupun rahang sebelah kanan. Berikut adalah tahapan utama yang dilakukan untuk mengukur lebar *cortical bone*.
 1. *Image Enchantment*
Sebelum diolah, kualitas citra ROI awal akan diperbaiki agar perhitungan dapat dilakukan dengan lebih mudah. Proses yang dilakukan pada tahapan ini meliputi *Thresholding* dan *High-pass Filtering*.
 2. *Cortical Margin Identification*
Pada tahapan ini akan dilakukan identifikasi batasan luar dari *cortical bone* dari citra yang dihasilkan pada

tahapan *Image Enchantment*. Kemudian dipilih *gradient* terbaik pada kedua batasan untuk mengukur lebar *cortical bone*.

3. *Distance Measurement*

Tahapan akhir yang dilakukan adalah menghitung lebar *cortical bone* dari batasan yang telah ditentukan pada tahapan *Cortical Margin Identification* menggunakan rumus jarak *Euclidean*.

Kedua tahapan *preprocessing* citra tersebut dilakukan dengan menggunakan algoritma dan *source code* yang terdapat pada penelitian yang telah dilakukan oleh Arifin dkk (2006, 2010)

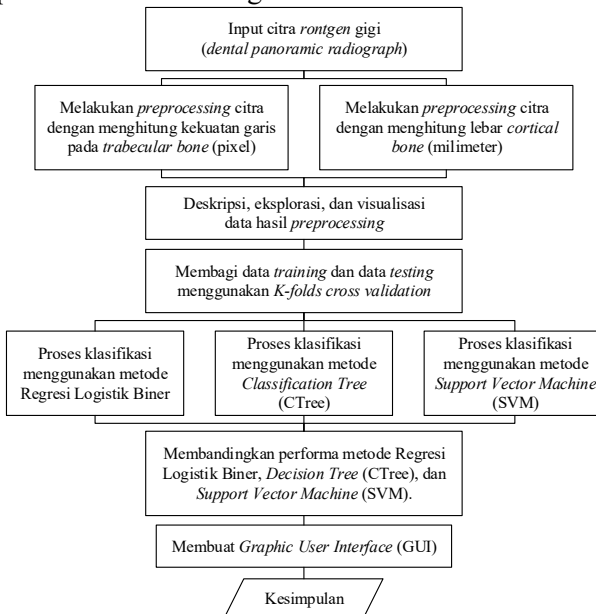
2. Mendeskripsikan secara umum data hasil *preprocessing* yang telah dilakukan secara visual.
3. Membagi data *training* dan data *testing* dengan menggunakan *K-folds cross validation* dengan *K* sebanyak 10, yaitu membagi banyak data yang ada menjadi data *training* dan data *testing* dengan perbandingan 90:10, dimana pembagian dilakukan sebanyak 10 kali, sehingga setiap data memiliki kesempatan untuk menjadi data *testing*.
4. Melakukan klasifikasi dengan menggunakan metode Regresi Logistik Biner. Klasifikasi dilakukan sebanyak 10 kali dengan menggunakan data *training* yang telah didapat pada proses *K-folds cross validation*. Selanjutnya akan dihitung nilai rata-rata dari semua akurasi yang didapat sebagai performa untuk metode tersebut.
5. Melakukan klasifikasi dengan menggunakan metode *Classification Tree* (CTree). Klasifikasi dilakukan sebanyak 10 kali dengan menggunakan data *training* yang telah didapat pada proses *K-folds cross validation*. Selanjutnya akan dihitung nilai rata-rata dari semua akurasi yang didapat sebagai performa untuk metode tersebut.
6. Melakukan klasifikasi dengan menggunakan metode *Support Vector Machine* (SVM). Klasifikasi dilakukan sebanyak 10 kali dengan menggunakan data *training* yang telah didapat pada proses *K-folds cross validation*. Selanjutnya akan dihi-

tung nilai rata-rata dari semua akurasi yang didapat sebagai performa untuk metode tersebut.

7. Mengulangi langkah 3, 4, dan 5 menggunakan variabel yang signifikan yang didapat pada langkah ke-3.
8. Membandingkan performa dari metode Regresi Logistik Biner, *Classification Tree* (CTree) dan *Support Vector Machine* (SVM) yang telah didapat baik untuk model yang menggunakan semua variabel, maupun model yang hanya menggunakan variabel yang berpengaruh secara signifikan. Ukuran performa yang digunakan sebagai pembanding adalah *accuracy*, *sensititivity*, dan *specivicity*.
9. Membuat program atau *Graphic User Interface* (GUI) dengan menggunakan metode terbaik yang telah didapat.

3.4 Diagram Alir

Diagram alir untuk langkah-langkah yang akan dilakukan dalam penelitian adalah sebagai berikut.



Gambar 3.1 Langkah-langkah yang akan dilakukan dalam penelitian.

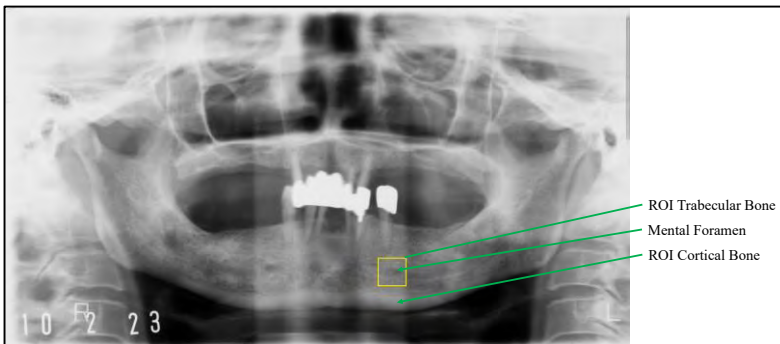
BAB IV

ANALISIS DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai karakteristik data tulang yang telah didapat pada tahapan *preprocessing*. Kemudian akan dilanjutkan dengan melakukan klasifikasi menggunakan metode Regresi Logistik Biner, *Classification Tree* (CTree) dan *Support Vector Machine* (SVM). Serta penjelasan mengenai *Graphic User Interface* (GUI) yang telah dibuat.

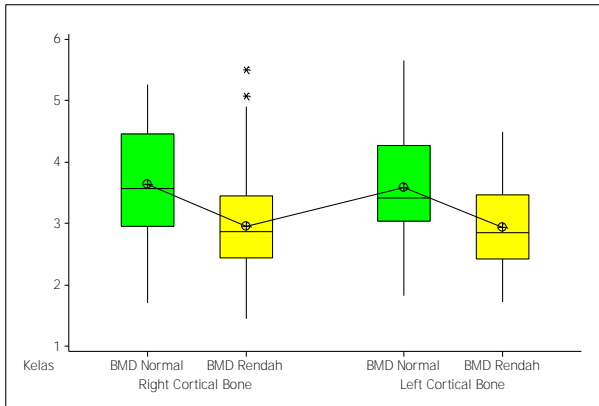
4.1 Deskripsi Data Hasil *Preprocessing* Citra Dental

Tahapan *preprocessing* dilakukan dengan menggunakan *source code* yang telah dibuat oleh Arifin dkk (2006, 2010) baik pada pengambilan data untuk bagian *cortical bone* maupun bagian *trabecular bone*. ROI (*Region Of Interest*) atau sampel bagian tulang yang dianalisis merupakan tulang yang berada pada bagian *mental foramen* atau bagian terbuka pada tulang rahang, dimana sampel *trabecular bone* yang dianalisis adalah bagian *trabecular bone* pada *mental foramen*, dan sampel *cortical bone* yang dianalisis adalah *cortical bone* di bawah *mental foramen*. Berikut ilustrasi ROI yang dianalisis dari citra rontgen gigi.

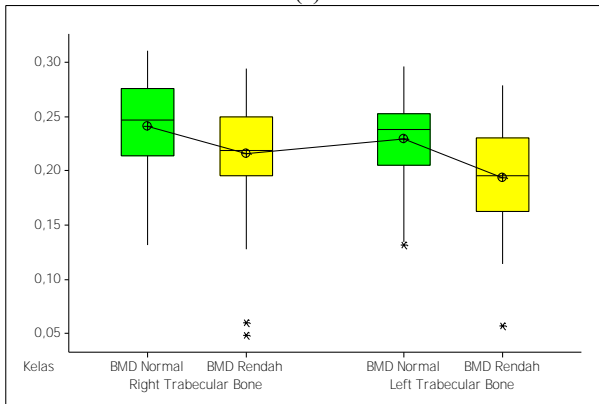


Gambar 4.1 Ilustrasi ROI yang dianalisis dalam *preprocessing*.

Setelah semua citra diproses, yaitu menghitung lebar *cortical bone* (mm) dan kekuatan garis pada *trabecular bone* (pixel), akan dilakukan deskripsi data secara visual menggunakan *boxplot*, dan didapat hasil sebagai berikut.



(a)



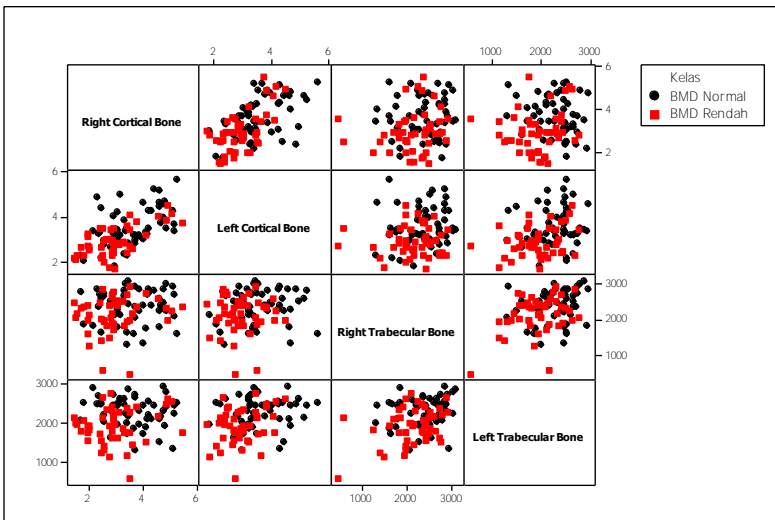
(b)

Gambar 4.2 Perbedaan sebaran data antara wanita dengan ‘BMD Normal’ dan wanita dengan ‘BMD Rendah’ pada (a) *cortical bone* dan (b) *trabecular bone*.

Melalui Gambar 4.2(a) dapat disimpulkan bahwa terdapat perbedaan sebaran data lebar *cortical bone* antara wanita dengan ‘BMD Normal’ dan wanita dengan ‘BMD Rendah’, dimana wanita dengan ‘BMD Normal’ cenderung memiliki lebar *cortical bone* (dalam mm) yang lebih lebar daripada wanita dengan kategori ‘BMD Rendah’, baik pada rahang sebelah kiri maupun sebelah kanan. Hal tersebut juga terjadi pada *trabecular bone*, dimana pada

Gambar 4.2(b) dapat dilihat bahwa wanita dengan ‘BMD Normal’ cenderung memiliki kekuatan garis pada *trabecular bone* (pixel) yang lebih banyak daripada wanita dengan ‘BMD Rendah’.

Kemudian akan dibuat *scatterplot* antar setiap variabel prediktor, yaitu lebar *cortical bone* pada sisi kiri maupun kanan dan kekuatan garis pada *trabecular bone* pada sisi kiri dan kanan, untuk melihat apakah sebaran data antar kategori dapat terpisah secara linier (*linearly separable*) atau tidak. Berikut adalah *scatterplot* yang didapat.



Gambar 4.3 Sebaran data antar kategori ‘BMD Normal’ dan ‘BMD Rendah’ pada setiap variabel prediktor.

Dapat diketahui bahwa secara umum sebaran data antara wanita dengan kategori ‘BMD Normal’ dan ‘BMD Rendah’ tidak *linearly separable*, dimana terdapat beberapa data wanita dengan kategori ‘BMD Normal’ berada pada daerah wanita dengan ‘BMD Rendah’ dan juga terjadi sebaliknya, dimana kasus ini disebut *linearly non-separable data*. Hal ini mengindikasikan bahwa pada kasus data ini sebaiknya menggunakan metode klasifikasi non-linier.

4.2 Klasifikasi BMD pada Wanita *Postmenopausal*

Setelah didapat data lebar *cortical bone* dan kekuatan garis pada *trabecular bone*, selanjutnya akan dilakukan pembuatan model untuk mengklasifikasikan wanita *postmenopausal* dengan ‘BMD Normal’ atau ‘BMD Rendah’. Metode klasifikasi yang digunakan adalah metode Regresi Logistik Biner, *Classification Tree* (CTree) dan *Support Vector Machine* (SVM). Berikut pembahasan lebih lanjut untuk masing-masing metode.

4.2.1 Klasifikasi dengan Metode Regresi Logistik Biner

Metode pertama yang akan digunakan dalam klasifikasi adalah metode Regresi Logistik Biner. Asumsi yang perlu dipenuhi dalam metode ini adalah tidak terjadi kasus multikolinieritas, dimana pengecekan asumsi ini dapat dilakukan menggunakan nilai VIF. Nilai VIF yang didapat untuk setiap variabel prediktor dapat dilihat pada Tabel 4.1 berikut.

Tabel 4.1 Nilai VIF setiap variabel prediktor

Variabel	Keterangan	VIF
RCB	<i>Right Cortical Bone</i>	1,654
LCB	<i>Left Cortical Bone</i>	1,785
RTB	<i>Right Trabecular Bone</i>	1,180
LTB	<i>Left Trabecular Bone</i>	1,285

Pada Tabel 4.1 dapat diketahui bahwa nilai VIF yang didapat, baik untuk variabel lebar *cortical bone* pada rahang sebelah kiri, variabel lebar *cortical bone* pada rahang sebelah kanan, variabel kekuatan garis pada *trabecular bone* pada rahang sebelah kiri, maupun variabel kekuatan garis pada *trabecular bone* pada rahang sebelah kanan lebih kecil dari 10, sehingga dapat disimpulkan bahwa pada data tidak terjadi kasus multikolinieritas.

Setelah pengecekan asumsi multikolinieritas terpenuhi, akan dilakukan pembuatan model dengan menggunakan metode *K-folds Cross Validation* (KCV) dengan nilai $K = 10$. Variabel yang digunakan dalam model adalah variabel RCB, LCB, RTB, dan LTB. Berikut adalah rincian hasil *accuracy*, *sensitivity*, dan *specificity* yang didapat pada setiap *fold*.

Tabel 4.2 Peforma metode Regresi Logistik Biner menggunakan semua variabel

<i>Folds</i>	<i>Accuracy</i>		<i>Sensitivity</i>		<i>Specificity</i>	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
1	71,11	80,00	68,29	60,00	73,47	100,00
2	74,44	70,00	70,73	80,00	77,55	60,00
3	72,22	80,00	68,29	80,00	75,51	80,00
4	75,56	60,00	75,61	40,00	75,51	80,00
5	71,11	90,00	68,29	80,00	73,47	100,00
6	74,44	70,00	70,73	80,00	77,55	60,00
7	72,22	100,00	68,29	100,00	75,51	100,00
8	74,44	40,00	70,73	60,00	77,55	20,00
9	71,11	90,00	68,29	100,00	73,47	80,00
10	73,33	80,00	75,56	0,00	71,11	88,89
<i>Mean</i>	73,00	76,00	70,48	68,00	75,07	76,89

Dapat diketahui bahwa model terbaik yang didapat adalah model pada *fold* ke-7, dimana *accuracy*, *sensitivity*, dan *specificity* yang didapat untuk data *testing* adalah 100%, dimana nilai ini lebih besar daripada nilai yang didapat pada model pada *fold* lainnya. Berikut adalah koefisien, statistik hitung, dan *pvalue* yang didapat pada *fold* ke-7.

Tabel 4.3 Nilai koefisien, statistik hitung, dan *pvalue* dari model pada *fold* ke-7 untuk semua variabel masuk dalam model

Variabel	Koefisien	Statistik hitung	<i>pvalue</i>
Konstan	6,749	3,751	0,000
RCB	-0,492	-1,555	0,120
LCB	-0,339	-0,845	0,398
RTB	-0,000	-0,989	0,323
LTB	-0,001	-2,226	0,026

Pada model terbaik tersebut masih terdapat beberapa variabel yang tidak berpengaruh signifikan pada alfa 5%. Maka akan dilakukan *backward elimination* untuk memilih variabel yang berpengaruh signifikan dengan cara mengeluarkan variabel yang paling tidak signifikan secara bertahap. Hasil akhir yang didapat dari *backward elimination* adalah sebagai berikut.

Tabel 4.4 Nilai koefisien, thitung, dan *pvalue* hasil *backward elimination*

Variabel	Koefisien	thitung	<i>pvalue</i>
Konstan	5,686	3,755	0,000
RCB	-0,687	-2,717	0,007
LTB	-0,002	-2,869	0,004

Didapat bahwa variabel yang berpengaruh signifikan dalam mengelompokkan wanita dengan ‘BMD Normal’ atau ‘BMD Rendah’ adalah variabel RCB dan LTb. Maka akan dilakukan tahapan *K-folds Cross Validation* kembali dengan menggunakan variabel RCB dan LTb. Berikut adalah hasil *accuracy*, *sensitivity*, dan *specificity* yang didapat untuk setiap *fold*.

Tabel 4.5 Peforma metode Regresi Logistik Biner menggunakan variabel RCB dan LTb

<i>Folds</i>	<i>Accuracy</i>		<i>Sensitivity</i>		<i>Specificity</i>	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
1	73,33	60,00	68,29	60,00	77,55	60,00
2	68,89	100,00	63,41	100,00	73,47	100,00
3	71,11	80,00	65,85	60,00	75,51	100,00
4	72,22	80,00	65,85	80,00	77,55	80,00
5	74,44	40,00	68,29	40,00	79,59	40,00
6	78,89	50,00	70,73	40,00	85,71	60,00
7	71,11	100,00	63,41	100,00	77,55	100,00
8	71,11	70,00	65,85	60,00	75,51	80,00
9	76,67	70,00	68,29	60,00	83,67	80,00
10	73,33	60,00	71,11	100,00	75,56	55,56
<i>Mean</i>	73,11	71,00	67,11	70,00	78,17	75,56

Setelah analisis dilakukan kembali dengan menggunakan variabel RCB dan LTb, didapat bahwa model terbaik adalah model yang dihasilkan pada *fold* ke-7, dimana hasil koefisien, statistik hitung, dan *pvalue* yang dihasilkan adalah sebagai berikut.

Tabel 4.6 Nilai koefisien, statistik hitung, dan *pvalue* dari model pada *fold* ke-7 dengan variabel prediktor RCB dan LTb

Variabel	Koefisien	Statistik hitung	<i>pvalue</i>
Konstan	5,703	3,721	0,000
RCB	-0,670	-2,671	0,008
LTB	-0,002	-3,051	0,002

Berdasarkan hasil pada Tabel 4.6, maka dapat dibuat model regresi logistik sebagai berikut.

$$\ln\left(\frac{\pi(X)}{1 - \pi(X)}\right) = 5,703 - 0,670.RCB - 0,002.LTB$$

Model yang telah didapat tersebut memberikan informasi bahwa jika lebar *cortical bone* pada rahang sebelah kanan bertambah satu milimeter, maka peluang wanita tersebut memiliki ‘BMD Rendah’ akan berkurang, dimana *odds ratio* wanita tersebut memiliki ‘BMD Rendah’ menjadi sebesar $\exp(-0,670) = 0,512$ kali, atau *odds ratio* wanita tersebut memiliki ‘BMD Normal’ menjadi sebesar $0,512^{-1} = 1,955$ kali. Informasi lain yang bisa diambil dari model tersebut adalah jika kekuatan garis pada *trabecular bone* pada rahang sebelah kiri pada citra bertambah satu pixel, maka peluang wanita tersebut memiliki ‘BMD Rendah’ akan berkurang, dimana *odds ratio* wanita tersebut memiliki ‘BMD Rendah’ menjadi sebesar $\exp(-0,002) = 0,998$ kali, atau *odds ratio* wanita tersebut memiliki ‘BMD Normal’ menjadi $0,998^{-1} = 1,002$ kali.

4.2.2 Klasifikasi dengan Metode *Classification Tree*

Metode kedua yang digunakan adalah metode *Classification Tree* yang merupakan metode *white box*. Berdasarkan hasil variabel yang signifikan yang telah didapat pada metode Regresi Logistik Biner, maka pada metode ini akan dilakukan dua jenis analisis, yaitu model yang menggunakan semua variabel dan model yang hanya menggunakan variabel RCB dan LTb. *Accuracy*, *sensitivity*, dan *specificity* yang didapat menggunakan semua variabel pada setiap *fold* dapat dilihat pada Tabel 4.7 berikut.

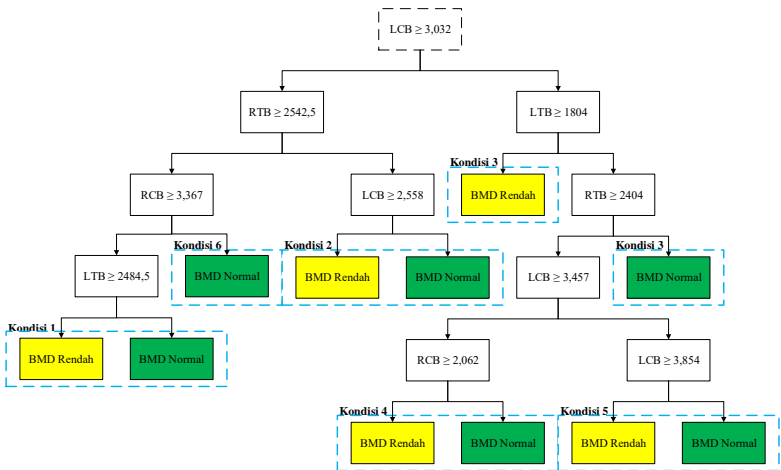
Tabel 4.7 Peforma metode *Classification Tree* menggunakan semua variabel

<i>Folds</i>	<i>Accuracy</i>		<i>Sensitivity</i>		<i>Specificity</i>	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
1	87,78	40,00	78,05	20,00	95,92	60,00
2	87,78	70,00	87,80	60,00	87,76	80,00
3	85,56	100,00	82,93	100,00	87,76	100,00
4	86,67	70,00	85,37	100,00	87,76	40,00
5	91,11	40,00	90,24	20,00	91,84	60,00
6	87,78	70,00	82,93	80,00	91,84	60,00
7	87,78	50,00	78,05	20,00	95,92	80,00

Tabel 4.7 Lanjutan

8	84,44	70,00	82,93	60,00	85,71	80,00
9	88,89	60,00	82,93	60,00	93,88	60,00
10	86,67	80,00	91,11	0,00	82,22	88,89
Mean	73,00	76,00	70,48	68,00	75,07	76,89

Pohon yang terbaik yang didapat dari 10 *fold* yang telah dilakukan adalah pohon pada *fold* ke-3, karena memiliki *accuracy*, *sensitivity*, dan *specificity* pada data *testing* yang lebih besar daripada pohon pada *fold* lainnya. Pohon yang dihasilkan pada *fold* ke-3 dapat dilihat pada Gambar 4.4 berikut.



Gambar 4.4 Pohon klasifikasi yang didapat pada *fold* ke-3 menggunakan semua variabel prediktor.

Beberapa pengelompokkan yang didapat berdasarkan pohon yang tersebut adalah sebagai berikut.

Tabel 4.8 Beberapa kondisi pengelompokkan hasil pohon klasifikasi dengan semua variabel masuk dalam model

Kondisi	Kategori	
	BMD Normal	BMD Rendah
1	$RCB < 3,367$	$RCB < 3,367$
	$LCB < 3,032$	$LCB < 3,032$
	$RTB < 2542,5$	$RTB < 2542,5$
	$LTB > 2484,5$	$LTB < 2484,5$

Tabel 4.8 Lanjutan

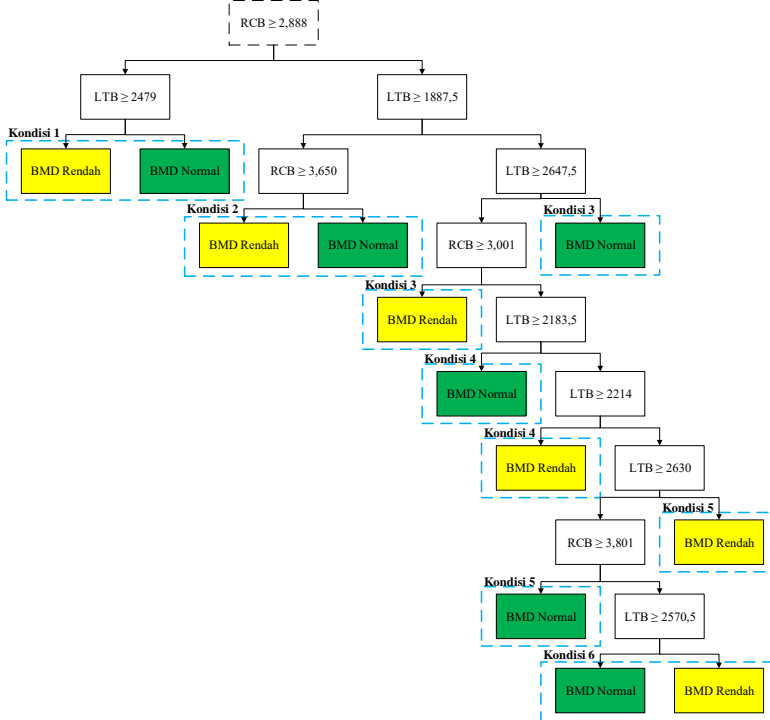
2	$2,558 < \textbf{LCB} < 3,032$	$\textbf{LCB} < 2,558$
	$\textbf{RTB} > 2542,5$	$\textbf{RTB} > 2542,5$
3	$\textbf{LCB} > 3,032$	$\textbf{LCB} > 3,032$
	$\textbf{RTB} > 2404$	$\textbf{LTB} < 1804$
	$\textbf{LTB} > 1804$	
4	$\textbf{RCB} > 2,062$	$\textbf{RCB} < 2,062$
	$3,032 < \textbf{LCB} < 3,457$	$3,032 < \textbf{LCB} < 3,457$
	$\textbf{RTB} < 2404$	$\textbf{RTB} < 2404$
	$\textbf{LTB} > 1804$	$\textbf{LTB} > 1804$
5	$\textbf{LCB} > 3,854$	$3,457 < \textbf{LCB} < 3,854$
	$\textbf{RTB} < 2404$	$\textbf{RTB} < 2404$
	$\textbf{LTB} > 1804$	$\textbf{LTB} > 1804$
6	$\textbf{RCB} > 3,367$	
	$\textbf{LCB} < 3,032$	
	$\textbf{RTB} < 2542,5$	

Kemudian analisis kedua yang akan dilakukan adalah membuat pohon klasifikasi hanya menggunakan variabel yang berpengaruh signifikan menurut metode Regresi Logistik Biner, yaitu variabel RCB dan LTB. Berikut adalah hasil *accuracy*, *sensitivity*, dan *specificity* yang didapat untuk setiap *fold*.

Tabel 4.9 Peforma metode *Classification Tree* hanya menggunakan variabel RCB dan LTB

<i>Folds</i>	<i>Accuracy</i>		<i>Sensitivity</i>		<i>Specificity</i>	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
1	88,89	60,00	80,49	60,00	95,92	60,00
2	82,22	100,00	78,05	100,00	85,71	100,00
3	87,78	60,00	80,49	60,00	93,88	60,00
4	87,78	70,00	82,93	40,00	91,84	100,00
5	87,78	90,00	87,80	80,00	87,76	100,00
6	87,78	60,00	82,93	40,00	91,84	80,00
7	91,11	60,00	90,24	60,00	91,84	60,00
8	85,56	70,00	78,05	60,00	91,84	80,00
9	88,89	80,00	80,49	100,00	95,92	60,00
10	86,67	90,00	84,44	100,00	88,89	88,89
<i>Mean</i>	87,44	74,00	82,59	70,00	91,54	78,89

Diketahui bahwa pohon yang dihasilkan pada *fold* ke-2 memiliki *accuracy*, *sensitivity*, dan *specificity* untuk data *testing* yang lebih besar daripada pohon yang dihasilkan pada *fold* lainnya, sehingga pohon ini merupakan pohon terbaik. Berikut adalah bentuk visual dari pohon yang dihasilkan pada *fold* ke-2 tersebut.



Gambar 4.5 Pohon klasifikasi yang didapat pada *fold* ke-2 hanya menggunakan variabel RCB dan LTB.

Berdasarkan pohon klasifikasi yang telah didapat tersebut, maka dapat dibuat beberapa kondisi pengelompokkan untuk mengelompokkan apakah seorang wanita *postmenopausal* dikategorikan memiliki ‘BMD Normal’ atau ‘BMD Rendah’ dengan melihat lebar *cortical bone* pada rahang sebelah kanan dan kekuatan garis pada *trabecular bone* pada rahang sebelah kiri. Kondisi pengelompokkan tersebut dapat dilihat pada Tabel 4.10 berikut.

Tabel 4.10 Beberapa kondisi pengelompokkan hasil pohon klasifikasi menggunakan variabel RCB dan LTB

Kondisi	Kategori	
	BMD Normal	BMD Rendah
1	<i>RCB</i> < 2,888	<i>RCB</i> < 2,888
	<i>LTB</i> > 2479	<i>LTB</i> < 2479
2	<i>RCB</i> > 3,650	2,888 < <i>RCB</i> < 3,650
	<i>LTB</i> < 1887,5	<i>LTB</i> < 1887,5
3	<i>RCB</i> > 2,888	2,888 < <i>RCB</i> < 3,001
	<i>LTB</i> > 2647,5	1887,5 < <i>LTB</i> < 2647,5
4	<i>RCB</i> > 3,001	<i>RCB</i> > 3,001
	1887,5 < <i>LTB</i> < 2183,5	2183,5 < <i>LTB</i> < 2214
5	3,001 < <i>RCB</i> < 3,801	<i>RCB</i> > 3,001
	2214 < <i>LTB</i> < 2630	2630 < <i>LTB</i> < 2647,5
6	<i>RCB</i> > 3,801	<i>RCB</i> > 3,801
	2214 < <i>LTB</i> < 2570,5	2570,5 < <i>LTB</i> < 2630

Pada Tabel 4.10 dapat diketahui bahwa jika seorang wanita memiliki lebar *cortical bone* pada rahang sebelah kanan lebih kecil dari 2,888 milimeter dan memiliki kekuatan garis pada *trabecular bone* pada rahang sebelah kiri lebih besar dari 2479 pixel, maka wanita tersebut dikategorikan memiliki ‘BMD Normal’. Contoh pengelompokkan lainnya adalah jika wanita tersebut memiliki lebar *cortical bone* pada rahang sebelah kanan lebih besar dari 2,888 milimeter dan memiliki kekuatan garis pada *trabecular bone* pada rahang sebelah kiri lebih besar dari 2647,5 pixel, maka wanita tersebut dikategorikan memiliki ‘BMD Normal’. Kemudian contoh kondisi untuk kategori ‘BMD Rendah’ adalah jika wanita tersebut memiliki lebar *cortical bone* pada rahang sebelah kanan lebih kecil dari 2,888 milimeter dan memiliki kekuatan garis pada *trabecular bone* pada rahang kiri lebih kecil dari 2479 pixel, maka wanita tersebut akan dikategorikan memiliki ‘BMD Rendah’. Contoh lainnya adalah jika wanita tersebut memiliki lebar *cortical bone* pada rahang sebelah kanan antara 2,888 milimeter dan 3,650 milimeter dan memiliki kekuatan garis pada *trabecular bone* pada rahang sebelah kiri lebih kecil dari 1887,5 pixel, maka wanita tersebut akan dikategorikan memiliki ‘BMD Rendah’.

4.2.3 Klasifikasi dengan Metode *Support Vector Machine*

Metode klasifikasi selanjutnya yang akan digunakan adalah salah satu metode klasifikasi non-linier, yaitu metode *Support Vector Machine*. Pada saat analisis menggunakan metode ini, digunakan prinsip *grid search* dalam mencari kombinasi parameter yang akan digunakan pada model. Prinsip *grid search* yang digunakan mengacu pada penelitian yang telah dilakukan oleh Hsu, Chang, dan Lin (2008), yaitu mencobakan kombinasi parameter dengan pertambahan yang bersifat eksponensial. Sebagai contoh pada fungsi kernel *radial basis*, akan dicobakan parameter *cost* sebesar $2^{-5}, 2^{-3}, \dots, 2^{15}$ dan parameter *gamma* sebesar $2^{-15}, 2^{-13}, \dots, 2^3$. Selanjutnya parameter yang memberikan performa terbaik akan dicobakan lagi kombinasi parameter dengan pertambahan yang lebih kecil.

Pada metode ini juga akan dilakukan dua jenis analisis, yaitu analisis menggunakan semua variabel, dan analisis hanya menggunakan variabel RCB dan LTB. Berikut adalah hasil *K-folds Cross Validation* yang didapat jika menggunakan semua variabel dan hanya menggunakan variabel RCB dan LTB.

Tabel 4.11 Rangkuman hasil KCV dari metode SVM untuk setiap fungsi kernel

Variabel dalam model	Kernel Parameter		<i>Linear</i>	<i>Polynomial</i>	<i>RBF</i>
	<i>Cost</i>		- 26	2 47	$2^{1,1}$ $2^{2,8}$
RCB	<i>Accuracy</i>	<i>Training</i>	74,22	77,00	74,33
		<i>Testing</i>	70,00	67,00	76,00
LCB	<i>Sensitivity</i>	<i>Training</i>	76,07	72,96	75,12
RTB		<i>Testing</i>	64,00	64,00	80,00
LTB	<i>Specificity</i>	<i>Training</i>	72,60	80,45	73,46
		<i>Testing</i>	69,78	71,78	74,67
RCB	<i>Accuracy</i>	<i>Training</i>	72,67	75,11	76,78
		<i>Testing</i>	73,00	75,00	77,00
LCB	<i>Sensitivity</i>	<i>Training</i>	70,30	64,00	67,66
LTB		<i>Testing</i>	74,00	68,00	70,00
	<i>Specificity</i>	<i>Training</i>	74,68	84,58	84,58
		<i>Testing</i>	73,78	83,78	85,87

Diketahui bahwa hasil yang didapat dengan menggunakan fungsi kernel *radial basis* dengan parameter *gamma* sebesar $2^{1,1}$ dan parameter *cost* sebesar $2^{2,8}$ merupakan hasil yang terbaik karena memiliki nilai *accuracy* dan *specificity* yang paling besar pada data *testing*, baik untuk model yang menggunakan semua variabel maupun model yang hanya menggunakan variabel RCB dan LTB. Namun model SVM yang hanya menggunakan variabel RCB dan LTB lebih baik daripada model SVM yang menggunakan semua variabel karena memiliki *accuracy testing* dan *specificity testing* yang lebih besar. Sehingga didapat bahwa model SVM yang terbaik adalah model dengan menggunakan fungsi kernel *radial basis* dengan parameter *gamma* sebesar $2^{1,1}$ dan parameter *cost* sebesar $2^{2,8}$ yang hanya menggunakan variabel RCB dan LTB. Rincian *accuracy*, *sensitivity*, dan *specificity* yang didapat dari model SVM terbaik untuk setiap *fold* dapat dilihat pada Tabel 4.12 berikut.

Tabel 4.12 Peforma metode *Support Vector Machine* menggunakan variabel RCB dan LTB

<i>Folds</i>	<i>Accuracy</i>		<i>Sensitivity</i>		<i>Specificity</i>	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
1	76,67	70,00	68,29	60,00	83,67	80,00
2	76,67	90,00	68,29	100,00	83,67	80,00
3	74,44	100,00	63,41	100,00	83,67	100,00
4	77,78	60,00	70,73	20,00	83,67	100,00
5	76,67	70,00	65,85	60,00	85,71	80,00
6	78,89	80,00	70,73	80,00	85,71	80,00
7	76,67	70,00	68,29	40,00	83,67	100,00
8	75,56	80,00	68,29	60,00	81,63	100,00
9	78,89	70,00	68,29	80,00	87,76	60,00
10	75,56	80,00	64,44	100,00	86,67	77,78
<i>Mean</i>	76,78	77,00	67,66	70,00	84,59	85,78

Pada Tabel 4.12 dapat diketahui bahwa model yang terbaik adalah model yang didapat pada *fold* ke-3, karena memiliki *accuracy*, *sensitivity*, dan *specificity* untuk data *testing* yang lebih besar daripada *fold* lainnya. Fungsi *hyperplane* yang didapat dari model pada *fold* ke-3 tersebut adalah

$$f(\mathbf{x}) = \sum_{m=1}^{85} \alpha_m \cdot K(\mathbf{SV}_m, \mathbf{x}) + 0,242$$

dimana fungsi kernel yang digunakan adalah fungsi kernel *radial basis* dengan parameter *gamma* sebesar $2^{1,1}$, dengan rumus

$$\begin{aligned} K(\mathbf{SV}, \mathbf{x}) &= \exp\left(-\frac{(\mathbf{SV} - \mathbf{x})(\mathbf{SV} - \mathbf{x})^T}{2\sigma^2}\right) \\ &= \exp\left(-\frac{(\mathbf{SV} - \mathbf{x})(\mathbf{SV} - \mathbf{x})^T}{2(2^{1,1})^2}\right) \\ &= \exp\left(-\frac{(\mathbf{SV} - \mathbf{x})(\mathbf{SV} - \mathbf{x})^T}{2^{3,2}}\right) \end{aligned}$$

sehingga fungsi *hyperplane* yang didapat menjadi

$$f(\mathbf{x}) = \sum_{m=1}^{85} \alpha_m \cdot \exp\left(-\frac{(\mathbf{SV}_m - \mathbf{x})(\mathbf{SV}_m - \mathbf{x})^T}{2^{3,2}}\right) + 0,242$$

dengan 0,242 merupakan nilai *bias*, α merupakan vektor koefisien dari *support vector* yang berukuran $(85; 1)$, \mathbf{SV} adalah matriks *support vector* yang berukuran $(85; 2)$, dan \mathbf{x} adalah vektor input yang akan diklasifikasi yang berukuran $(1; 2)$, yaitu lebar *cortical bone* pada rahang sebelah kanan dan kekuatan garis pada *trabecular bone* pada rahang sebelah kiri dari seorang wanita *postmenopausal* yang akan dicek apakah termasuk memiliki ‘BMD Normal’ atau ‘BMD Rendah’. Jika didapat nilai $f(\mathbf{x}) \geq 0$ maka wanita dengan data tersebut akan dikategorikan memiliki ‘BMD Normal’, namun jika nilai $f(\mathbf{x}) \leq 0$ maka wanita tersebut akan dikategorikan memiliki ‘BMD Rendah’.

4.2.4 Perbandingan Performa Antar Metode

Setelah didapat nilai-nilai *accuracy*, *sensitivity*, dan *specificity* dari metode Regresi Logistik Biner, *Classification Tree*, dan *Support Vector Machine* baik ketika menggunakan semua variabel maupun hanya menggunakan variabel RCB dan LTB, maka akan dilakukan perbandingan metode berdasarkan nilai-nilai tersebut

untuk memilih metode yang terbaik yang digunakan pada GUI. Rangkuman hasil performa untuk setiap metode yang digunakan dapat dilihat pada Tabel 4.13 berikut.

Tabel 4.13 Rangkuman performa dari metode Regresi Logistik Biner (RLB), *Classification Tree* (CTree), dan *Support Vector Machine* (SVM)

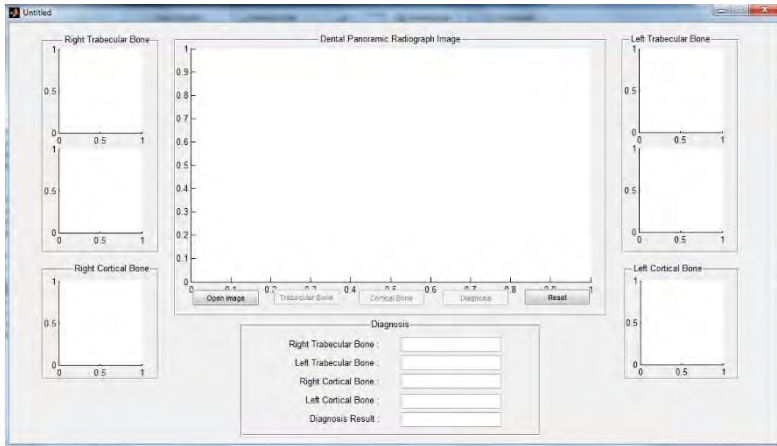
Variabel pada model	Metode	<i>Accuracy</i>		<i>Sensitivity</i>		<i>Specificity</i>	
		<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>
Semua Variabel	RLB	73,00	76,00	70,48	68,00	75,07	76,89
	CTree	73,00	76,00	70,48	68,00	75,07	76,89
	SVM	74,33	76,00	75,12	80,00	73,46	74,67
RCB, LTB	RLB	73,11	71,00	67,11	70,00	78,17	75,56
	CTree	87,44	74,00	82,59	70,00	91,54	78,89
	SVM	76,78	77,00	67,66	70,00	84,59	85,78

Pada Tabel 4.13 dapat diketahui bahwa metode *Classification Tree* (CTree) yang hanya menggunakan variabel RCB dan LTB memiliki nilai *accuracy*, *sensitivity*, dan *specificity* untuk data *training* paling besar. Namun untuk data *testing*, metode *Support Vector Machine* (SVM) yang hanya menggunakan variabel RCB dan LTB memiliki nilai *accuracy* dan *specificity* yang lebih besar daripada metode CTree dan memiliki nilai *sensitivity* yang sama dengan metode CTree. Metode yang akan dipilih adalah metode yang memiliki performa yang baik pada data *testing*, sehingga metode yang dipilih dan akan digunakan pada GUI adalah metode SVM dengan variabel yang digunakan adalah RCB dan LTB.

4.3 *Graphic User Interface* (GUI) Deteksi Dini Osteoporosis dari Citra Rontgen Gigi

Setelah diketahui bahwa metode terbaik adalah metode *Support Vector Machine* (SVM) menggunakan fungsi kernel *radial basis* dengan parameter *gamma* bernilai $2^{1,1}$, maka akan dibuat suatu *Graphic User Interface* (GUI) untuk membuat sistem otomatis dalam deteksi dini osteoporosis, yaitu mengklasifikasikan apakah seorang wanita *postmenopausal* memiliki BMD yang normal atau rendah melalui DPR. GUI yang dibuat merupakan modifikasi dan penggabungan dari *source code* yang telah dibuat oleh Arifin dkk (2006) untuk mengukur lebar *cortical bone*, dan

source code yang telah dibuat oleh Arifin dkk (2010) untuk mengukur kekuatan garis pada *trabecular bone*. Tampilan awal dari GUI yang dibuat adalah sebagai berikut.



Gambar 4.6 Tampilan menu utama GUI pada kondisi awal.

Pada menu utama GUI tersebut terdapat 6 panel, yaitu 1 panel untuk menampilkan citra asli, 2 panel untuk *trabecular bone* dimana pada masing-masing panel akan menampilkan citra ROI awal dan hasil pengolahan, 2 panel untuk menampilkan hasil pengukuran lebar *cortical bone*, dan 1 panel untuk menampilkan hasil pengukuran yang telah dilakukan serta diagnosis kategori BMD dari wanita *postmenopausal* tersebut. Terdapat 5 tombol yang digunakan dalam GUI tersebut, yaitu tombol ‘*Open Image*’, ‘*Trabecular Bone*’, ‘*Cortical Bone*’, ‘*Diagnosis*’, dan ‘*Reset*’. Penjelasan untuk masing-masing tombol tersebut dapat dilihat pada Tabel 4.14 berikut.

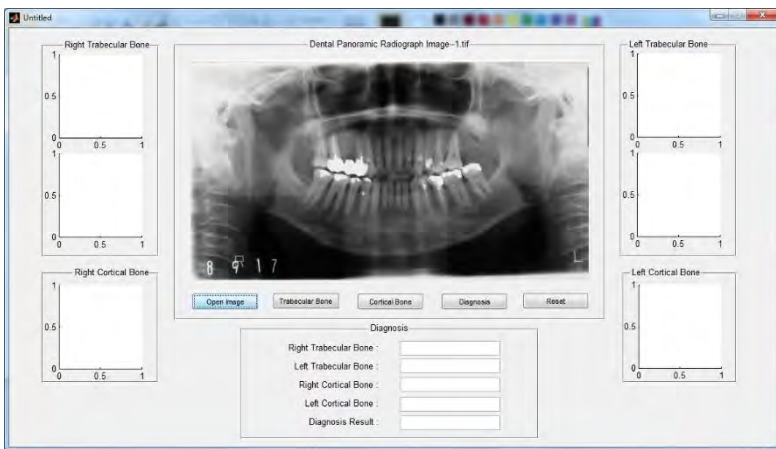
Tabel 4.14 Penjelasan fungsi tombol-tombol pada GUI

Nama Tombol	Fungsi
‘ <i>Open Image</i> ’	Memilih dan menampilkan citra yang akan dianalisis.
‘ <i>Trabecular Bone</i> ’	Melakukan proses perhitungan kekuatan garis pada <i>trabecular bone</i> pada ROI.

Tabel 4.14 Lanjutan

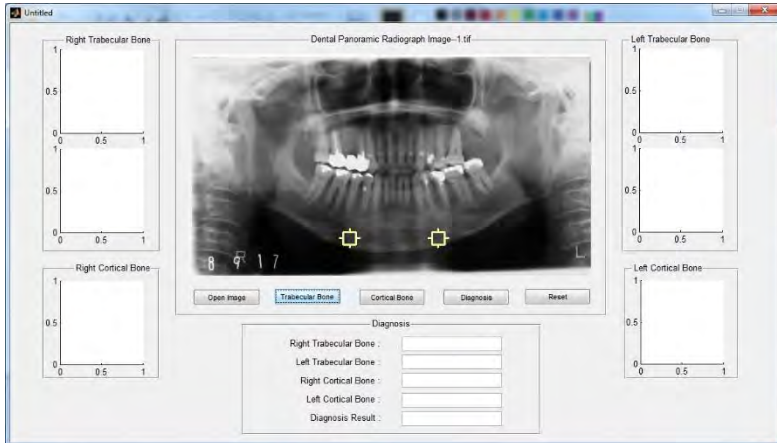
'Cortical Bone'	Melakukan proses perhitungan lebar <i>cortical bone</i> dibawah <i>mental foramen</i> .
'Diagnosis'	Mengklasifikasikan wanita <i>postmenopausal</i> yang sedang dianalisis apakah diklasifikasikan memiliki ' <i>Normal Bone Mineral Density</i> ' atau ' <i>Low Bone Mineral Density</i> '.
'Reset'	Menghapus semua proses yang telah dilakukan dan mengembalikan kondisi GUI menjadi kondisi awal.

Langkah pertama yang akan dilakukan adalah memilih DPR yang akan dianalisis menggunakan tombol '*Open Image*'. Setelah DPR dipilih, tampilan pada GUI akan menjadi seperti pada Gambar 4.7 berikut.

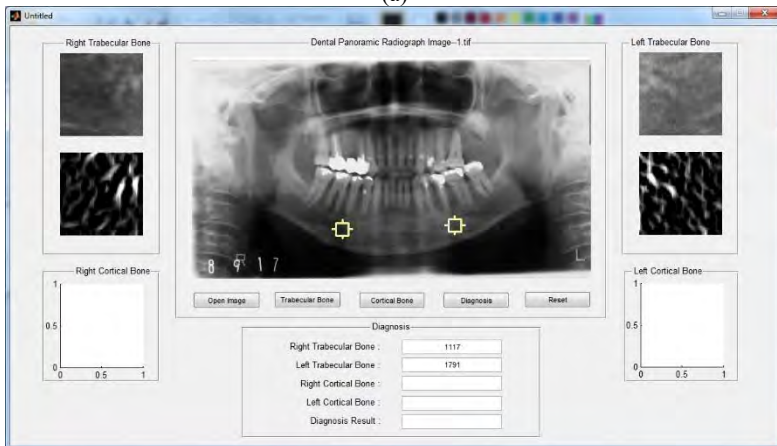
**Gambar 4.7** Tampilan GUI setelah citra rontgen gigi dipilih.

Setelah citra asli DPR dipilih dan keluar pada GUI, langkah selanjutnya adalah melakukan proses perhitungan kekuatan garis pada *trabecular bone*. Proses ini dapat dilakukan dengan menekan tombol '*Trabecular Bone*' sehingga muncul 2 kotak kuning pada GUI yang akan digunakan untuk memilih ROI yang akan dianalisis. Kedua kotak kuning tersebut kemudian dipindahkan oleh *User* sehingga kedua kotak kuning tersebut berada pada *mental foramen*. Kemudian program akan melakukan analisis pada kedua ROI yang

telah dipilih tersebut sehingga didapat citra hasil proses, yang akan ditampilkan pada masing-masing panel '*Trabecular Bone*', dan kekuatan garis pada *trabecular bone* dalam pixel yang akan ditampilkan pada panel '*Diagnosis*'. Tampilan GUI ketika tombol '*Trabecular Bone*' ditekan dan hasil akhirnya dapat dilihat pada Gambar 4.8 berikut.



(a)



(b)

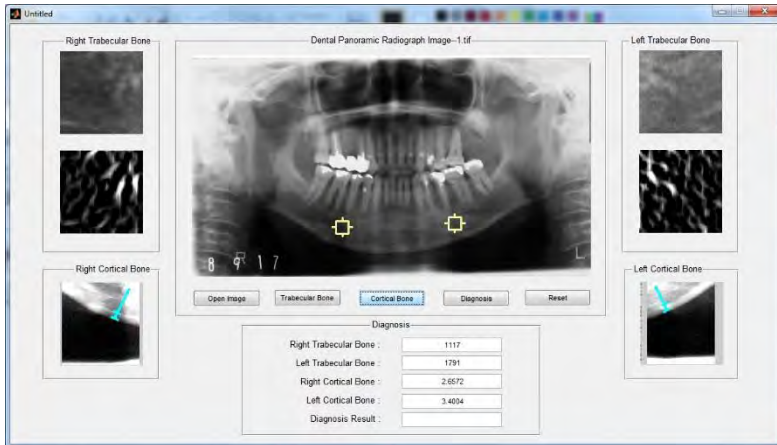
Gambar 4.8 Tampilan GUI (a) ketika tombol '*Trabecular Bone*' ditekan dan (b) setelah proses perhitungan pada *trabecular bone* selesai.

Tahapan selanjutnya adalah mengukur lebar *cortical bone* yang dapat dilakukan dengan menekan tombol '*Cortical Bone*'. Ketika tombol tersebut ditekan akan muncul citra hasil rotasi sebesar 90° searah jarum jam seperti pada Gambar 4.9 berikut.



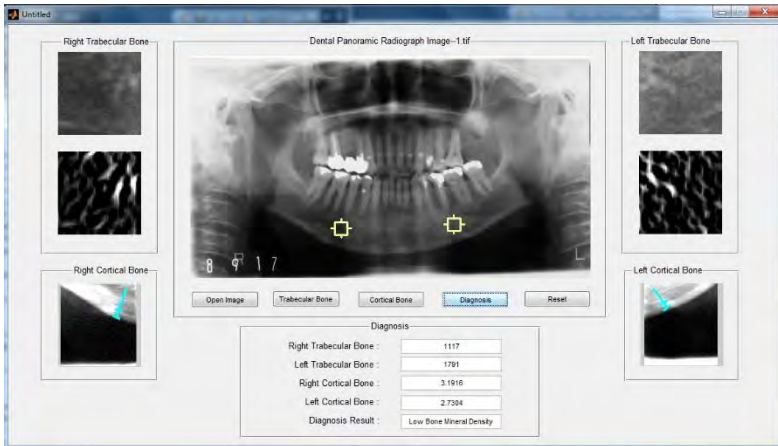
Gambar 4.9 Tampilan yang muncul ketika tombol '*Cortical Bone*' ditekan.

Pada citra hasil rotasi tersebut, *User* akan diminta untuk memilih dua lokasi dimana *mental foramen* pada rahang sebelah kiri dan rahang sebelah kanan berada. Setelah kedua lokasi *mental foramen* tersebut dipilih, citra hasil rotasi tersebut akan ditutup secara otomatis dan pada tampilan menu utama GUI, pada panel '*Cortical Bone*' akan muncul ROI *cortical bone* beserta ilustrasi hasil pengukuran lebarnya. Kemudian pada panel '*Diagnosis*' juga akan muncul hasil pengukuran lebar *cortical bone* yang telah didapat. Tampilan GUI setelah proses pada tombol '*Cortical Bone*' selesai dilakukan dapat dilihat pada Gambar 4.10 berikut.



Gambar 4.10 Tampilan GUI ketika proses pada tombol ‘Cortical Bone’ selesai.

Setelah data lebar *cortical bone* dan kekuatan garis pada *trabecular bone* pada rahang sebelah kiri dan rahang sebelah kanan didapat, selanjutnya akan dilakukan proses klasifikasi apakah wanita *postmenopausal* yang memiliki citra rontgen gigi tersebut memiliki BMD yang normal atau rendah. Klasifikasi dilakukan dengan menggunakan model terbaik yang telah didapat pada tahapan sebelumnya, yaitu metode *Support Vector Machine* (SVM) menggunakan fungsi kernel *radial basis* dengan parameter *gamma* bernilai $2^{1,1}$. Variabel yang digunakan dalam klasifikasi adalah variabel lebar *cortical bone* pada rahang sebelah kanan (dalam milimeter) dan kekuatan garis pada *trabecular bone* pada rahang sebelah kiri (dalam pixel). Proses tersebut dapat dilakukan dengan menekan tombol ‘Diagnosis’. Hasil klasifikasi yang telah diproses akan ditampilkan pada panel ‘Diagnosis’, dimana kemungkinan hasil yang terjadi adalah ‘Normal Bone Mineral Density’, berarti wanita tersebut didiagnosis memiliki BMD yang normal, dan ‘Low Bone Mineral Density’, berarti wanita tersebut didiagnosis memiliki BMD yang rendah, sehingga berpotensi terkena osteoporosis. Tampilan GUI setelah tombol ‘Diagnosis’ ditekan adalah sebagai berikut.



Gambar 4.11 Tampilan GUI ketika proses pada tombol ‘*Diagnosis*’ selesai.

Keseluruhan proses analisis telah selesai ketika hasil diagnosis telah keluar. Jika *User* ingin melakukan analisis menggunakan citra baru, *User* dapat menekan tombol ‘*Reset*’, sehingga tampilan GUI akan kembali menjadi tampilan awal.

(Halaman ini sengaja dikosongkan)

LAMPIRAN

Lampiran 1. Data Hasil *Preprocessing* dari Citra Rontgen Gigi

Citra	Kelas	RCB (mm)	LCB (mm)	RTB (pixel)	LTB (pixel)
1.tif	BMD Rendah	2,6518	3,0291	1850	2356
2.tif	BMD Rendah	1,9307	2,7107	1988	1752
3.tif	BMD Normal	5,1053	4,2872	2970	1341
4.tif	BMD Rendah	4,1279	3,1916	2778	1529
5.tif	BMD Normal	3,4475	2,8297	2480	2174
6.tif	BMD Normal	4,3114	4,2847	2434	2111
7.tif	BMD Normal	4,0116	4,6659	1342	2457
8.tif	BMD Rendah	1,7941	2,3418	2382	2091
9.tif	BMD Rendah	1,9982	2,5258	2148	1938
10.tif	BMD Normal	2,8960	3,2933	2769	2736
11.tif	BMD Normal	2,3295	4,8858	2895	2500
12.tif	BMD Rendah	1,5003	2,1436	2089	2003
13.tif	BMD Normal	4,6382	4,6636	1823	2129
14.tif	BMD Rendah	5,4949	3,7292	2391	1751
15.tif	BMD Rendah	3,0691	2,8461	2438	1613
16.tif	BMD Normal	1,7018	2,3110	2816	2071
17.tif	BMD Normal	3,8688	3,1916	2283	1981
18.tif	BMD Normal	2,9231	4,0392	2454	2474
19.tif	BMD Rendah	2,3902	3,5216	2517	1737
⋮	⋮	⋮	⋮	⋮	⋮
99.tif	BMD Rendah	2,8385	3,0006	2662	1816
100.tif	BMD Rendah	3,4287	3,6170	1959	1153

Keterangan

RCB : Lebar *cortical bone* pada rahang kanan.

LCB : Lebar *cortical bone* pada rahang kiri.

RTB : Kekuatan garis *trabecular bone* pada rahang kanan.

LTB : Kekuatan garis *trabecular bone* pada rahang kiri.

Lampiran 2. Nilai Estimasi α dan *Support Vectors* dari Model Terbaik SVM

i	α	<i>Support Vector</i>	
		RCB	LTB
1	-5.377	0.1101	0.0889
2	-3.0265	0.3579	1.3672
3	-3.2268	1.3202	0.0822
4	-8.8501	-0.7683	-0.1565
5	-4.7516	-0.8503	0.8563
6	-4.537	-1.5064	0.9009
7	-2.9741	-0.4245	1.3426
8	-5.6812	0.3353	-0.6183
9	-3.4327	0.9476	-0.0517
10	-8.115	-0.3651	-0.3461
11	-4.6942	1.8707	0.8742
12	-4.9736	-0.9737	0.8162
13	-3.1979	1.2373	0.0175
14	-3.015	-0.9558	1.2869
15	-3.9781	1.3907	1.8446
16	-3.4549	1.0526	0.7425
17	-7.4535	0.0604	-0.8994
18	-4.5725	-0.2851	0.7738
19	-1.2701	-1.1281	1.7732
20	-4.6287	-0.2477	0.7381
21	-7.7639	-0.2567	-0.3774
22	-3.6567	-0.1675	1.0727
23	-4.6934	-0.8687	0.8719
24	-4.3814	1.7969	0.718
25	-4.6401	0.6837	-0.8972
26	-4.0062	0.4942	0.4012

Lampiran 2. Nilai Estimasi α dan *Support Vectors* dari Model Terbaik SVM (Lanjutan)

i	α	<i>Support Vector</i>	
		RCB	LTB
27	-3.509	1.3892	0.4659
28	-4.8813	-0.9105	0.8318
29	-4.1807	1.4944	1.51
30	-3.7377	1.4816	0.6399
31	-4.0034	0.7697	-0.5313
32	-4.6353	0.5185	-0.3417
33	-4.7538	-0.3982	0.7582
34	-3.9572	1.7172	-1.7694
35	-3.7531	1.2312	-1.4058
36	-10.9438	-0.5246	-1.1403
37	-3.1802	1.2644	-0.0115
38	-3.9827	1.7989	0.3053
39	-7.5353	-0.7642	0.1714
40	-3.5654	0.657	0.7202
41	-5.1223	0.0512	0.3031
42	-4.1007	0.0026	0.8251
43	-3.4364	-0.1554	1.1597
44	-4.3096	-0.2567	0.8541
45	-7.8358	0.3514	-1.8386
46	-3.0229	1.4471	-0.4131
47	-3.5037	0.8792	0.225
48	-10.1592	-1.5822	-0.1409
49	-3.4139	-0.0268	1.1508
50	4.2411	-1.8342	0.0376
51	12.1338	-0.5184	1.452
52	6.467	-0.3982	-0.0584

Lampiran 2. Nilai Estimasi α dan *Support Vectors* dari Model Terbaik SVM (Lanjutan)

i	α	<i>Support Vector</i>	
		RCB	LTB
53	10.0636	1.6954	0.9857
54	0.0644	-0.846	-1.292
55	0.8675	-1.3105	-0.6272
56	1.6779	-1.7522	-0.3684
57	11.4035	1.2192	0.1313
58	8.126	-0.4547	0.4124
59	10.7029	1.4694	0.805
60	10.6432	0.411	0.6421
61	3.4252	0.0919	-2.1888
62	6.2217	0.2772	-1.5419
63	9.3448	-0.4401	0.7068
64	9.7265	0.1957	0.37
65	6.8931	-0.4245	0.08
66	4.4776	-0.5484	-0.4198
67	7.1001	0.1661	-0.857
68	3.492	-1.4927	-0.0963
69	1.0966	0.1967	-3.4917
70	3.7885	-0.4802	-0.7098
71	10.3592	1.5264	1.1396
72	11.3686	-0.7958	1.2289
73	9.988	0.0696	0.6221
74	5.6489	-0.3604	-0.335
75	8.6051	-0.2218	0.3811
76	2.6217	-0.4547	-1.1381
77	8.0572	-0.6612	0.4949
78	2.1246	-1.7775	-0.2926

Lampiran 2. Nilai Estimasi α dan *Support Vectors* dari Model Terbaik SVM (Lanjutan)

i	α	<i>Support Vector</i>	
		RCB	LTB
79	5.5516	-0.818	0.051
80	10.691	2.0949	-0.8548
81	1.0732	-0.9148	-0.886
82	1.8484	-1.2948	-0.4376
83	6.1608	-0.0131	-0.8079
84	9.3502	0.7697	-1.35
85	6.4704	-0.0637	-0.5692

Lampiran 3. *Syntax* Klasifikasi Metode Regresi Logistik Biner dengan *K-folds Cross Validation* menggunakan *Software Matlab*

```

%-- Membaca Data
clear; clc;
[X Yraw] = xlsread('E:/DataOsteoJapan.xlsx');
Yraw = double(nominal(Yraw(2:length(Yraw),2)));
for i=1:length(Yraw)
    if Yraw(i)==2
        Y(i)=1;
    else
        Y(i)=2;
    end;
end;
Y = Y';

%-- RL dengan KCV
Y_Normal = find(Y==2); Y_Rendah = find(Y==1);
n1 = length(Y_Normal); n0 = length(Y_Rendah);
k1 = round(n1/10); k0 = round(n0/10);
sam_Normal = randsample(Y_Normal, n1, false);
sam_Rendah = randsample(Y_Rendah, n0, false);

```

Lampiran 3. *Syntax* Klasifikasi Metode Regresi Logistik Biner dengan *K-folds Cross Validation* menggunakan *Software Matlab* (Lanjutan)

```

for i=1:10
    %-- Membagi Data Menjadi Training & Testing
    if i==10
        Xtest = X([sam_Normal((9*k1+1):n1);
                    sam_Rendah((9*k0+1):n0)],:);
        Ytest = Y([sam_Normal((9*k1+1):n1);
                    sam_Rendah((9*k0+1):n0)]);
        Xtrain = X([sam_Normal(1:(9*k1));
                     sam_Rendah(1:(9*k0))],:);
        Ytrain = Y([sam_Normal(1:(9*k1));
                     sam_Rendah(1:(9*k0))]);
    elseif i==1
        Xtest = X([sam_Normal(1:k1);
                    sam_Rendah(1:k0)],:);
        Ytest = Y([sam_Normal(1:k1);
                    sam_Rendah(1:k0)]);
        Xtrain = X([sam_Normal((k1+1):n1);
                     sam_Rendah((k0+1):n0)],:);
        Ytrain = Y([sam_Normal((k1+1):n1);
                     sam_Rendah((k0+1):n0)]);
    else
        Xtest = X([sam_Normal(((i-1)*k1+1):i*k1);
                    sam_Rendah(((i-1)*k0+1):i*k0)],:);
        Ytest = Y([sam_Normal(((i-1)*k1+1):i*k1);
                    sam_Rendah(((i-1)*k0+1):i*k0)]);
        Xtrain = X([sam_Normal(1:((i-1)*k1);
                           sam_Normal((i*k1+1):n1);
                           sam_Rendah(1:((i-1)*k0));
                           sam_Rendah((i*k0+1):n0)],:);
        Ytrain = Y([sam_Normal(1:((i-1)*k1);
                           sam_Normal((i*k1+1):n1);
                           sam_Rendah(1:((i-1)*k0));
                           sam_Rendah((i*k0+1):n0)]);

    end;

    %-- Model RL
    Model_RL_KCV{i} = mnrfit(Xtrain,Ytrain);

    %-- Training
    gy = (Model_RL_KCV{i}(2:3)*Xtrain'
          +Model_RL_KCV{i}(1));
    prob = exp(gy)./(1+exp(gy));
    Y_Fits = round(prob);

```

Lampiran 3. *Syntax* Klasifikasi Metode Regresi Logistik Biner dengan *K-folds Cross Validation* menggunakan *Software Matlab* (Lanjutan)

```

for j=1:length(prob)
    if prob(j)>0.5
        Y_Fits(j)=1;
    else
        Y_Fits(j)=2;
    end;
end;
CfM = confusionmat(Ytrain, Y_Fits);
Akurasi_RL_Train(i) = sum(diag(CfM))/sum(sum(CfM))
                    *100;
Sens_RL_Train(i) = CfM(1,1)/sum(CfM(1,:))*100;
Spec_RL_Train(i) = CfM(2,2)/sum(CfM(2,:))*100;

%-- Testing
gy = (Model_RL_KCV{i}(2:3)'*Xtest'
      +Model_RL_KCV{i}(1));
prob = exp(gy)./(1+exp(gy));
Y_Fits = round(prob);
for j=1:length(prob)
    if prob(j)>0.5
        Y_Fits(j)=1;
    else
        Y_Fits(j)=2;
    end;
end;
CfM = confusionmat(Ytest, Y_Fits);
Akurasi_RL_Test(i) = sum(diag(CfM))/sum(sum(CfM))
                  *100;
Sens_RL_Test(i) = CfM(1,1)/sum(CfM(1,:))*100;
Spec_RL_Test(i) = CfM(2,2)/sum(CfM(2,:))*100;
end;

Akurasi_RL_KCV = [mean(Akurasi_RL_Train)
                  mean(Akurasi_RL_Test)]
Sens_RL_KCV = [mean(Sens_RL_Train) mean(Sens_RL_Test)]
Spec_RL_KCV = [mean(Spec_RL_Train) mean(Spec_RL_Test)]

[Akurasi_RL_Train' Akurasi_RL_Test' Sens_RL_Train'
 Sens_RL_Test' Spec_RL_Train' Spec_RL_Test']

```

Lampiran 4. *Syntax* Klasifikasi Metode *Classification Tree* dengan *K-folds Cross Validation* menggunakan *Software Matlab*

```
%-- Membaca Data
clear; clc;
[X Y] = xlsread('E:/DataOsteoJapan.xlsx');
Y = strncmp(Y(2:length(Y),2), 'BMD Normal', 10);

%-- CTree KCV
Y_Normal = find(Y==1); Y_Rendah = find(Y==0);
n1 = length(Y_Normal); n0 = length(Y_Rendah);
k1 = round(n1/10); k0 = round(n0/10);
sam_Normal = randsample(Y_Normal, n1, false);
sam_Rendah = randsample(Y_Rendah, n0, false);

for i=1:10
    %-- Membagi Data Menjadi Training & Testing
    if i==10
        Xtest = X([sam_Normal((9*k1+1):n1);
                    sam_Rendah((9*k0+1):n0)],:);
        Ytest = Y([sam_Normal((9*k1+1):n1);
                    sam_Rendah((9*k0+1):n0)]);
        Xtrain = X([sam_Normal(1:(9*k1));
                     sam_Rendah(1:(9*k0))],:);
        Ytrain = Y([sam_Normal(1:(9*k1));
                     sam_Rendah(1:(9*k0))]);
    elseif i==1
        Xtest = X([sam_Normal(1:k1);
                    sam_Rendah(1:k0)],:);
        Ytest = Y([sam_Normal(1:k1);
                    sam_Rendah(1:k0)]);
        Xtrain = X([sam_Normal((k1+1):n1);
                     sam_Rendah((k0+1):n0)],:);
        Ytrain = Y([sam_Normal((k1+1):n1);
                     sam_Rendah((k0+1):n0)]);
    else
        Xtest = X([sam_Normal(((i-1)*k1+1):i*k1);
                    sam_Rendah(((i-1)*k0+1):i*k0)],:);
        Ytest = Y([sam_Normal(((i-1)*k1+1):i*k1);
                    sam_Rendah(((i-1)*k0+1):i*k0)]);
        Xtrain = X([sam_Normal(1:((i-1)*k1));
                     sam_Normal((i*k1+1):n1);
                     sam_Rendah(1:((i-1)*k0));
                     sam_Rendah((i*k0+1):n0)],:);
    end
end
```

Lampiran 4. *Syntax* Klasifikasi Metode *Classification Tree* dengan *K-folds Cross Validation* menggunakan *Software Matlab* (Lanjutan)

```

        Ytrain = Y([sam_Normal(1:((i-1)*k1));
                    sam_Normal((i*k1+1):n1);
                    sam_Rendah(1:((i-1)*k0));
                    sam_Rendah((i*k0+1):n0)]);

    end;

    %-- Model CTree
    Model_CTree_KCV{i} =
        ClassificationTree.fit(Xtrain,Ytrain);

    %-- Training
    Y_Fits = predict(Model_CTree_KCV{i}, Xtrain);
    CfM = confusionmat(Ytrain, Y_Fits);
    Akurasi_CTree_Train(i) = sum(diag(CfM))/sum(sum(CfM))
        *100;
    Sens_CTree_Train(i) = CfM(1,1)/sum(CfM(1,:))*100;
    Spec_CTree_Train(i) = CfM(2,2)/sum(CfM(2,:))*100;

    %-- Testing
    Y_Fits = predict(Model_CTree_KCV{i}, Xtest);
    CfM = confusionmat(Ytest, Y_Fits);
    Akurasi_CTree_Test(i) = sum(diag(CfM))/sum(sum(CfM))
        *100;
    Sens_CTree_Test(i) = CfM(1,1)/sum(CfM(1,:))*100;
    Spec_CTree_Test(i) = CfM(2,2)/sum(CfM(2,:))*100;

end;

Akurasi_CTree_KCV = [mean(Akurasi_CTree_Train)
                    mean(Akurasi_CTree_Test)]
Sens_CTree_KCV = [mean(Sens_CTree_Train)
                 mean(Sens_CTree_Test)]
Spec_CTree_KCV = [mean(Spec_CTree_Train)
                 mean(Spec_CTree_Test)]

[Akurasi_CTree_Train' Akurasi_CTree_Test'
 Sens_CTree_Train' Sens_CTree_Test'
 Spec_CTree_Train' Spec_CTree_Test']

```

Lampiran 5. *Syntax* Klasifikasi Metode *Support Vector Machine* dengan *K-folds Cross Validation* menggunakan *Software Matlab*

```
%-- Membaca Data
clear; clc;
[X Y] = xlsread('E:/DataOsteoJapan.xlsx');
Y = strncmp(Y(2:length(Y),2), 'BMD Normal', 10);

%-- SVM KCV
vek_cost = 2.^1.1;
vek_sigma = 2.^2.8;

Y_Normal = find(Y==1); Y_Rendah = find(Y==0);
n1 = length(Y_Normal); n0 = length(Y_Rendah);
k1 = round(n1/10); k0 = round(n0/10);
sam_Normal = randsample(Y_Normal, n1, false);
sam_Rendah = randsample(Y_Rendah, n0, false);

for i=1:10
    %-- Membagi Data Menjadi Training & Testing
    if i==10
        Xtest = X([sam_Normal((9*k1+1):n1);
                    sam_Rendah((9*k0+1):n0)],:);
        Ytest = Y([sam_Normal((9*k1+1):n1);
                    sam_Rendah((9*k0+1):n0)]);
        Xtrain = X([sam_Normal(1:(9*k1));
                     sam_Rendah(1:(9*k0))],:);
        Ytrain = Y([sam_Normal(1:(9*k1));
                     sam_Rendah(1:(9*k0))]);
    elseif i==1
        Xtest = X([sam_Normal(1:k1);
                    sam_Rendah(1:k0)],:);
        Ytest = Y([sam_Normal(1:k1);
                    sam_Rendah(1:k0)]);
        Xtrain = X([sam_Normal((k1+1):n1);
                     sam_Rendah((k0+1):n0)],:);
        Ytrain = Y([sam_Normal((k1+1):n1);
                     sam_Rendah((k0+1):n0)]);
    else
        Xtest = X([sam_Normal(((i-1)*k1+1):i*k1);
                    sam_Rendah(((i-1)*k0+1):i*k0)],:);
        Ytest = Y([sam_Normal(((i-1)*k1+1):i*k1);
                    sam_Rendah(((i-1)*k0+1):i*k0)]);
    end
end
```

Lampiran 5. Syntax Klasifikasi Metode *Support Vector Machine* dengan *K-folds Cross Validation* menggunakan *Software Matlab* (Lanjutan)

```

Xtrain = X([sam_Normal(1:((i-1)*k1));
            sam_Normal((i*k1+1):n1);
            sam_Rendah(1:((i-1)*k0));
            sam_Rendah((i*k0+1):n0)],:);
Ytrain = Y([sam_Normal(1:((i-1)*k1));
            sam_Normal((i*k1+1):n1);
            sam_Rendah(1:((i-1)*k0));
            sam_Rendah((i*k0+1):n0)]);

end;
%-- Model SVM
options = optimset('maxiter', 100000);
Model_SVM_KCV{i} = svmtrain(Xtrain,Ytrain,
                             'boxconstraint',vek_cost,
                             'kernel_function','rbf',
                             'rbf_sigma',vek_sigma,
                             'method','QP',
                             'quadprog_opts',options);

%-- Training
Y_Fits = svmclassify(Model_SVM_KCV{i}, Xtrain);
CfM = confusionmat(Ytrain, Y_Fits);
Akurasi_SVM_Train(i) = sum(diag(CfM))/sum(sum(CfM))
                    *100;
Sens_SVM_Train(i) = CfM(1,1)/sum(CfM(1,:))*100;
Spec_SVM_Train(i) = CfM(2,2)/sum(CfM(2,:))*100;

%-- Testing
Y_Fits = svmclassify(Model_SVM_KCV{i}, Xtest);
CfM = confusionmat(Ytest, Y_Fits);
Akurasi_SVM_Test(i) = sum(diag(CfM))/sum(sum(CfM))
                    *100;
Sens_SVM_Test(i) = CfM(1,1)/sum(CfM(1,:))*100;
Spec_SVM_Test(i) = CfM(2,2)/sum(CfM(2,:))*100;

end;

Akurasi_SVM_KCV = [mean(Akurasi_SVM_Train)
                  mean(Akurasi_SVM_Test)]
Sens_SVM_KCV = [mean(Sens_SVM_Train) mean(Sens_SVM_Test)]
Spec_SVM_KCV = [mean(Spec_SVM_Train) mean(Spec_SVM_Test)]

[Akurasi_SVM_Train' Akurasi_SVM_Test' Sens_SVM_Train'
 Sens_SVM_Test' Spec_SVM_Train' Spec_SVM_Test']

```


Lampiran 6. Surat Pernyataan Legalitas Data**SURAT PERNYATAAN**

Saya yang bertanda tangan di bawah ini, mahasiswa Jurusan Statistika FMIPA ITS :

Nama : M. Rizky Fauzy

NRP : 1312 100 053

menyatakan bahwa data yang digunakan dalam Tugas Akhir ini merupakan data sekunder yang diambil dari penelitian yaitu :

Sumber : Jurnal *Osteoporosis International*, Vol. 17 tahun 2006, halaman 753-759, dengan judul : *Computer-aided System for Measuring the Mandibular Cortical Width on Dental Panoramic Radiograph in Identifying Postmenopausal Women with Low Bone Mineral Density*.

Keterangan : Citra *Dental Panoramic Radiograph* dari 100 wanita *postmenopausal* yang berusia lebih dari 50 tahun.


Surat Pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Mengetahui

Pembimbing Tugas Akhir

Surabaya, 29 Juni 2016


Dr. Agus Zainal Arifin, S.Kom, M.Kom
NIP. 19720809 199512 1 001


M. Rizky Fauzy
NRP. 1312 100 053

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Data Hasil <i>Preprocessing</i> dari Citra Rontgen Gigi	65
Lampiran 2 Nilai Estimasi α dan <i>Support Vectors</i> dari Model Terbaik SVM	66
Lampiran 3 <i>Syntax</i> Klasifikasi Metode Regresi Logistik Biner dengan <i>K-folds Cross Validation</i> Menggunakan <i>Software Matlab</i>	69
Lampiran 4 <i>Syntax</i> Klasifikasi Metode <i>Classification Tree</i> dengan <i>K-folds Cross Validation</i> menggunakan <i>Software Matlab</i>	72
Lampiran 5 <i>Syntax</i> Klasifikasi Metode <i>Support Vector Machine</i> dengan <i>K-folds Cross Validation</i> Menggunakan <i>Software Matlab</i>	74
Lampiran 6 Surat Pernyataan Legalitas Data	76

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan yang dapat ditarik berdasarkan hasil analisis yang telah dilakukan adalah terdapat perbedaan lebar *cortical bone* dan kekuatan garis pada *trabecular bone* antara wanita *postmenopausal* dengan ‘BMD Normal’ dan wanita *postmenopausal* dengan ‘BMD Rendah’, baik pada rahang sebelah kiri maupun rahang sebelah kanan. Wanita *postmenopausal* yang memiliki ‘BMD Normal’ cenderung memiliki lebar *cortical bone* yang lebih lebar dan kekuatan garis pada *trabecular bone* yang lebih besar daripada wanita *postmenopausal* dengan ‘BMD Rendah’.

Kesimpulan lain yang didapat adalah metode *Support Vector Machine* (SVM) merupakan metode yang terbaik dibanding metode Regresi Logistik Biner dan metode *Classification Tree* dalam mengklasifikasikan wanita *postmenopausal*, apakah dikategorikan memiliki ‘BMD Normal’ ataupun ‘BMD Rendah’. Model SVM yang digunakan merupakan model dengan fungsi kernel *radial basis* dengan parameter *gamma* bernilai $2^{1,1}$. Variabel yang berpengaruh signifikan yang digunakan dalam model SVM adalah variabel RCB (*Right Cortical Bone*) dan LTB (*Left Trabecular Bone*), yang didapat dari hasil *backward elimination* pada model Regresi Logistik Biner. Model SVM tersebut kemudian diterapkan pada GUI bersamaan dengan program pengukuran lebar *cortical bone* dan kekuatan garis pada *trabecular bone*, sehingga didapat suatu GUI yang dapat melakukan deteksi dini wanita *postmenopausal* dari citra rontgen gigi (*Dental Panoramic Radiograph*) apakah didiagnosis memiliki BMD yang normal atau BMD yang rendah.

5.2 Saran

Berdasarkan analisis yang telah dilakukan pada penelitian ini, beberapa saran untuk penelitian selanjutnya adalah sebagai berikut.

1. Pada penelitian selanjutnya diharapkan sebelum dilakukan klasifikasi, dilakukan tahapan *preprocessing* terlebih dahulu

dengan cara mengatasi *noise* atau *outlier* yang ada pada data, sehingga didapat hasil klasifikasi yang lebih akurat.

2. Diharapkan pada penelitian selanjutnya menggunakan metode klasifikasi yang berbeda atau pengembangan dari metode yang telah digunakan pada tugas akhir ini, sehingga dapat dibandingkan dengan metode yang telah digunakan pada penelitian ini dan dipilih metode yang lebih baik.

DAFTAR PUSTAKA

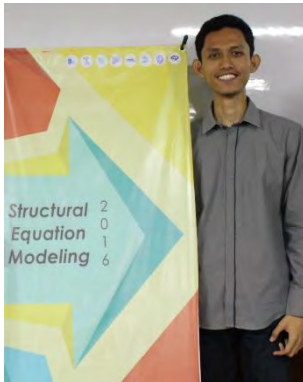
- Abidin, Z., & Arifin, A. Z. (2008). Analisis Kerapatan Trabecular Bone Berbasis Graph Berbobot pada Citra Panorama Gigi untuk Identifikasi Osteoporosis. *Jurnal Ilmiah Teknologi Informasi*, 7(2), 57–62. Retrieved from <http://juti.if.its.ac.id/index.php/juti/article/view/174/129>
- Arhandi, P. P., Arifin, A. Z., & Khotimah, W. N. (2014). Pengukuran Otomatis Lebar Cortical Bone pada Dental Panoramic Radiograph, 5(2), 455–467.
- Arifin, A. Z., Asano, A., Taguchi, A., Nakamoto, T., Ohtsuka, M., Tsuda, M., ... Tanimoto, K. (2006). Computer-aided system for measuring the mandibular cortical width on dental panoramic radiographs in identifying postmenopausal women with low bone mineral density. *Osteoporosis International*, 17(5), 753–759. doi:10.1007/s00198-005-0045-2
- Arifin, A. Z., Yuniarti, A., Dewi, L. R., Asano, A., Taguchi, A., Nakamoto, T., & Razak, A. (2010). Computer-aided diagnosis for osteoporosis based on trabecular bone analysis using panoramic radiographs. *Dental Jurnal Majalah Kedokteran Gigi*, 43(3).
- Assosiation, A. D. (2012). Glossary of Dental Clinical and Administrative Terms. Retrieved February 12, 2015, from <http://www.ada.org/en/publications/cdt/glossary-of-dental-clinical-and-administrative-ter>
- Auria, L., & Moro, R. A. (2008). Support Vector Machines (SVM) as a Technique for Solvency Analysis. *DIW Berlin Discussion Paper*, 811(August), 1–16. Retrieved from <http://ssrn.com/abstract=1424949>
- Draper, N., & Smith, H. (1992). *Analisis Regresi Terapan*. Jakarta: PT. Gramedia Pustaka Utama.
- Fachruddin, M. I. (2015). *Perbandingan Metode Random Forest Classification dan Support Vector Machine untuk Deteksi Epilepsi Menggunakan Data Rekaman Electroencephalograph (EEG)*. Institut Teknologi Sepuluh

Nopember.

- Gokgoz, E., & Subasi, A. (2015). Comparison of Decision Tree Algorithms for EMG Signal Classification Using DWT. *Biomedical Signal Processing and Control*, 18, 138–144. doi:10.1016/j.bspc.2014.12.005
- Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2009). *Digital Image Processing Using MATLAB* (2nd ed.). USA: Gatesmark Publishing.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques*. (J. Kacprzyk & L. C. Jain, Eds.) (12th ed.). Berlin: Springer.
- Group, W. S. (2003). *Prevention and Management of Osteoporosis*. (Who Scientific Group, Ed.) *World Health Organization Technical Report Series* (Vol. 921). Geneva: World Health Organization. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15293701>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (3rd ed.). USA: Morgan Kaufmann.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley & Sons, Inc.
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2008). A Practical Guide to Support Vector Classification. *BJU International*, 101(1), 1396–400. doi:10.1177/02632760022050997
- Lian, J., Gorski, J., & Ott, S. (2004). Bone Structure and Function. *American Society for Bone and Mineral Research: Bone Curriculum*. Retrieved February 13, 2016, from <https://depts.washington.edu/bonebio/ASBMRed/structure.html>
- Ogut, J. O., Piepho, H.-P., & Schulz-Streeck, T. (2011). A Comparison of Random Forests, Boosting and Support Vector Machines for Genomic Selection. *BMC Proceedings*, 5 Suppl 3(Suppl 3), S11. Retrieved from <http://www.biomedcentral.com/1753-6561/5/S3/S11>
- Pangkalan Ide. (2012). *Agar Tulang Sehat* (1st ed.). Jakarta: PT. Elex Media Komputindo.

- Setiawan, & Kusriani, D. E. (2010). *Ekonometrika*. (N. W. K., Ed.). Yogyakarta: CV. ANDI OFFSET.
- Taguchi, A., Suei, Y., Sanada, M., Ohtsuka, M., Nakamoto, T., Sumida, H., ... Tanimoto, K. (2004). Validation of dental panoramic radiography measures for identifying postmenopausal women with spinal osteoporosis. *American Journal of Roentgenology*, 183(6), 1755–1760. doi:10.2214/ajr.183.6.01831755
- Tribun News. (2012). Dua dari Lima Orang Indonesia Berisiko Osteoporosis. Retrieved February 9, 2016, from <http://m.tribunnews.com/kesehatan/2012/10/12/dua-dari-lima-orang-indonesia-berisiko-osteoporosis>
- Tribun News. (2015). Jarang Olahraga? Hati-hati Osteoporosis Mengintai. Retrieved February 9, 2016, from <http://m.tribunnews.com/lifestyle/2015/11/02/jarang-olahraga-hati-hati-osteoporosis-mengintai>
- Wardhana, G. W., Arifin, A. Z., & Navastara, D. A. (2009). Pengukuran Lebar Cortical Bone Berbasis Active Shape Model (ASM).
- White, S. C., & Rudolph, D. J. (1999). Alterations of the trabecular pattern of the jaws in patients with osteoporosis. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, 88(5), 628–635. doi:S1079210499003030 [pii]
- Wihandika, R. C. (2013). *Deteksi Percabangan Trabecular Bone Menggunakan Multiscale Cosfire Filter pada Identifikasi Osteoporosis*. Institut Teknologi Sepuluh Nopember.

BIODATA PENULIS



Penulis bernama M. Rizky Fauzy atau biasa dipanggil Rizky lahir di Dumai, 23 Maret 1995. Penulis adalah anak kedua dari empat bersaudara yang lahir dari pasangan Suami Istri Asber Oktafianus dan Masdalena. Penulis telah menyelesaikan pendidikan di TK Muslimin, SD N Binaan Khusus Kota Dumai (2001-2007), SMP Negeri 2 Dumai (2007-2009), dan SMA N Plus Propinsi Riau (2009-2012). Kemudian penulis melanjutkan pendidikan kuliah S1 di Jurusan Statistika ITS melalui jalur SNMPTN Undangan dengan NRP 1312100053. Penulis pernah aktif dalam organisasi kampus sebagai *Staff Training Development* di Divisi Statistics Computer Course (SCC) HIMASTA-ITS periode 2013-2014, kemudian menjadi *Manager Training Development* Divisi SCC HIMASTA-ITS periode 2014-2015. Selain melakukan aktivitas di kampus, penulis juga pernah mengikuti beberapa kompetisi lomba seperti *Data Mining* GEMASTIK 8 (Finalis), *Disaster Data Challenge* (DDC) UII 2016 (Juara 3), *Data Mining* GEMASTIK 7 (Medali Perunggu), dan beberapa lomba lainnya. Penulis juga pernah menjadi pembicara pada pelatihan *Structural Equation Modeling* yang diadakan oleh Divisi SCC-HIMASTA ITS pada tahun 2016. Penulis pernah memiliki pengalaman magang selama satu setengah bulan di PT. Pertamina (Persero) di *Marketing Operation Region* (MOR) V. Segala kritik, saran, dan diskusi mengenai Tugas Akhir ini dapat disampaikan melalui email mrizkyfauzy23@gmail.com.